

Storm clouds over Brussels

Europe's research commissioner should develop a bolder vision for his 'European Research Area', while explicitly addressing the conflicting demands on the European Union's research programme.

In January 2000, just a few months after taking over as European Commissioner for Research, Philippe Busquin floated his big idea — the European Research Area (ERA). The goal was to create a coherent European science policy, including unfettered mobility of researchers and support at the European Union (EU) level for large-scale research infrastructure.

Eighteen months later, there are fears that Busquin may have overreached himself. Many observers cannot see how this ambitious goal will be met, particularly given the questions that now hang over the centre-piece of Busquin's plans, the sixth five-year Framework programme for research, due to start in 2003.

The document that launched the ERA noted that it could not be created through the Framework programmes alone. But the disquiet surrounding Busquin's proposal for the sixth Framework bodes ill for his wider vision. The plan will be discussed next week by ministers from the member states, chaired by Thomas Östros, Sweden's science minister. Ominously, Östros has talked of "major question marks".

Östros is worried about support for basic and long-term research, which he feels are undervalued in Busquin's proposal. He also doubts whether procedures for selecting projects — and for monitoring grant-holders — are strong enough to guarantee scientific quality.

Later this year, the proposal will be debated by the European Parliament, whose members have begun to call for numerous changes. Some want to strengthen non-nuclear energy research, others want to streamline application procedures further, while others desire a merger between the 'genomics and biotechnology for health' and the 'food safety and health risks' programmes.

Busquin's fundamental problem is that Framework serves a multitude of masters with divergent demands. Leading scientific EU nations, for instance, are pleased with plans to support fewer, larger projects and delegate much of the management to the labs involved. But the Eastern and Central European nations queuing up to join the EU — and who already participate in Framework — complain

that this will cut them out of the picture (see *Nature* **411**, 512; 2001). Successful integration of these nations, meanwhile, is a major component of the ERA proposal.

Although Busquin's task is not easy, he can be criticized for failing to balance long-term goals and political realities. His wholesale shift of EU funding into larger projects and support for infrastructure makes some sense in the context of the ERA. But many voices from the grassroots have made it clear that they want more continuity and the retention of some of the familiar small-scale research networks funded under the current Framework programme.

Busquin is also open to charges of vagueness. The ERA proposal, for instance, called for EU support for projects with "variable geometry" — bilateral and multilateral collaborations initiated by the member states. Such funding would be allowed under Busquin's sixth Framework proposal. But few policymakers or scientists have a clear idea of what this would mean in practice — and this uncertainty is breeding misgivings.

Although delays are likely, it is not too late for Busquin to work with Europe's research ministers and the parliament to make the sixth Framework more attractive to the best scientists across the continent. In parallel, he should begin planning the concrete moves needed to make the ERA a reality.

In the long run, however, the ERA is likely to remain a bloodless vision unless there is an independent, flexible and self-administered pan-European funding body which — unlike the ponderous Framework — can react quickly to unexpected scientific developments.

A possible starting point exists. The European Coal and Steel Community — which created a 'common market' that eventually became subsumed within the EU — will be converted next year into a research fund for the coal and steel industries, endowed with some 1.3 billion euros (US\$1.1 billion). If European law were changed to allow this body to be boosted with further private and public money, it might eventually be transformed into a true pan-European research foundation. ■

Some advice for the president

George W. Bush needs a scientific adviser even more than the US research community needs a voice in the White House.

The two issues that so deeply divided the United States and Europe during President George W. Bush's transatlantic visit last week have this much in common: for both global warming and missile defence, governments need to interpret complex scientific arguments before deciding on a particular course of action.

It is doubly unfortunate, therefore, that Bush went to Europe without a scientific adviser by his side — or even at the end of the phone. Some five months into his administration, the position of the president's special assistant on science and technology, who also runs the White House Office of Science and Technology Policy (OSTP), remains unfilled.

Bill Clinton took just four days to name Jack Gibbons as candidate for the position after his January 1993 inauguration. Bush cannot be judged by that standard: after all, the planning of his administration

was delayed by the lengthy dispute over his election last November. But with scientific issues now having more bearing than ever on practical politics, the current delay is eating away at the credibility of Bush's positions on missile defence, climate change, energy policy and stem-cell research. It is also eroding the prospect that the OSTP will ever become fully staffed and functional under the Bush administration: other branches of the White House are greedy for its staff slots, and may soon borrow some of them.

Meanwhile, Bush's critics in the US scientific community — who are not entirely without influence on public opinion — are left to fulminate from the outside over the president's various policy positions. The sooner Bush appoints a respected Republican scientist to head the OSTP, the better are his chances of avoiding embarrassment in forthcoming debates on these positions. ■





Flash flood
Tropical storm
devastates labs
in Texas
p874



German jitters
Ethics council
delays approval for
stem-cell research
p875



Pot looters
Archaeologists
express fears
over thieves
p876



US isolated?
Europe woos
Japan over
Kyoto protocol
p877

Database of molecular probes set to boost chemical genetics

David Adam

A powerful open-access database that would link proteins to their role in cells and pinpoint their effects on whole organisms is being planned by the US National Cancer Institute (NCI).

Under the plans, the NCI would ask scientists around the world to deposit data on the effects of small chemical molecules on proteins, cell pathways and tissue formation. An additional important aspect would be the compilation of data on how such molecules affect an organism's phenotype — the observable characteristics that result from the expression of its genes.

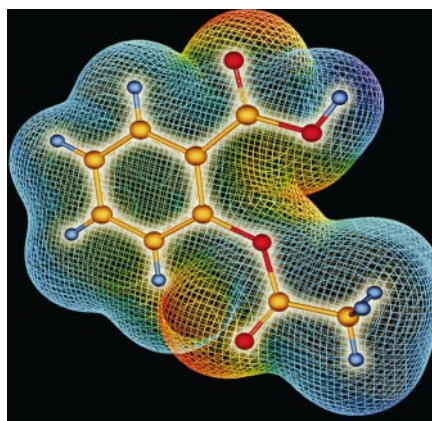
The database has been dubbed 'ChemBank' by some of its supporters, who see it as a chemical version of GenBank, the online repository for genetic data.

A search of ChemBank for all known effects of a given molecule could reveal connections between previously unrelated biological functions. If a particular molecule that is known to bind a protein, for example, is also found to disrupt a metabolic pathway, this could show that the protein is directly involved in that pathway. The repository would effectively hold all the available data in the emerging field of 'chemical genetics' (see *Nature* **407**, 282–284; 2000).

Plans for the database are at an early stage. But the NCI will take the first steps towards it later this year by establishing dedicated 'molecular targeting laboratories'. These labs will synthesize tens of thousands of small molecules and screen them for their biological effects. The information produced will be added to existing similar types of data to form the backbone of the database.

"This struck me as something that we really needed," says NCI director Richard Klausner. "If we're ever going to have an annotated database of molecular probes — chemical entities that recognize and may or may not do things to gene products — then it's going to require a pretty systematic effort to characterize and develop those probes."

Although Klausner believes that such probes will ultimately be used in cancer treatment, he insists that the new scheme is "not a drug-discovery programme". Rather,



Shapes of the future: the proposed database may uncover new uses even for familiar aspirin.

it aims to compile a complete catalogue of the effects of molecular probes on gene products, pathways and phenotypes, he says.

The NCI is currently considering proposals from potential hosts and will announce its support support for one or more

molecular targeting labs within a few months. Funding for the first year will be around \$10 million, and that will subsequently increase, says Klausner.

One candidate to host such a laboratory is Harvard University's Institute of Chemistry and Cell Biology. "It's a big, big undertaking and pretty speculative at the moment," says Stuart Schreiber, the institute's co-director. "But it could help us realize the full potential of chemical genetics." Existing collections of data on small molecules generally consist of information on protein binding and rarely include phenotypic effects, he notes. "Binding data by themselves don't illuminate biology."

Craig Crews, a cell biologist who works in chemical genetics at Yale University, says the database would be "extremely valuable". As well as revealing new aspects of cell biology, he adds, it will streamline efforts to relate research on new molecules to existing data. "Right now you have to be something of a detective to make those links," Crews says. ■

Asthma study death spurs inquiry

Paul Smaglik, Washington

The death this month of a woman taking part in an asthma study at Johns Hopkins University, Baltimore, has prompted a formal investigation by the US Office for Human Research Protections (OHRP).

The experiment, funded by the National Institutes of Health (NIH), was designed to study how lung function differs in asthmatic and healthy people. In an attempt to find out why the airways remain open when irritating chemicals are inhaled, healthy people were asked to inhale substances that affect these airways.

Some of the research subjects, who were paid about \$60 a visit up to a maximum of \$365, inhaled hexamethonium, which can temporarily paralyse some nerves in the airways. Those subjects, and others who had inhaled a saline-solution control, then took an asthma test.

The death at Johns Hopkins comes at a time when research using human subjects is under increased scrutiny in the United States. After the death in 1999 of a young man in a gene-therapy experiment at the University of Pennsylvania, the US government transferred responsibility for overseeing such research from the NIH's Office for Protection from Research Risks to the new OHRP, under the direct supervision of the US health secretary. Investigation of this latest death will provide the first major test of the new arrangement.

The exploratory focus of the asthma experiment has already attracted criticism. Vera Sharav, president of the New York-based Citizens for Responsible Care & Research, calls it a "fishing expedition" best reserved for animals, not people. "People should be protected from this kind of experiment," she says. ■

Texas facilities count cost of tropical storm

Rex Dalton, San Diego

A week after a tropical storm flooded research facilities in Houston, Texas, many scientists were still locked out of their labs, assessing a disaster that has wiped out years of studies on a variety of diseases.

More than 30,000 research animals — mostly mice and rats — were destroyed, an untold number of reagents obliterated and many investigations wrecked by the record rainfall from tropical storm Allison, which swept in off the Gulf of Mexico in the early morning of 9 June.

The rain flooded basements and knocked out power to research buildings and hospitals on the sprawling Texas Medical Center campus on the city's south side — a vast biomedical complex that attracts nearly \$1 billion of grant support each year from the US National Institutes of Health (NIH).

Research facilities at Baylor College of Medicine and the University of Texas Health Science Center at Houston were hit hardest by the storm. A week later, Baylor's eight research buildings were operating on a maintenance basis only, with researchers and



Calm after the storm: the floods in Houston washed away years of work at the city's laboratories.

students working to minimize long-term damage. The main University of Texas research building in Houston remained closed, along with the adjacent Memorial Hermann Hospital.

James Patrick, dean of research at Baylor, said that 30,000 of the institution's 130,000 research rodents have been destroyed. "A large portion of these were genetically modified mice," says Patrick.

"It is a major setback that reagents are lost," says Baylor geneticist Huda Zoghbi, "but the significant loss is time — all the work that will have to be repeated. Many of us collaborate with researchers everywhere. The impact of this will be worldwide."

On the afternoon after the flood, Katherina Walz, a postdoc working in the laboratory of geneticist James Lupski, found that more than two years of chromosomal-engineering experiments were destroyed when her colony of 150 mice drowned. "I just cried," says Walz, when she learned what happened.

Fortunately, Walz says, she discovered that three female mice — created with a genetic deficiency to mimic the Smith-Magenis mental-retardation disease — survived in another research building. It will take her more than six months to breed sufficient mice for experiments. "I feel like now I want to work more than I did before," she says.

Federal officials have declared the research region a disaster area, which means that special repair funding will be available. NIH officials are also waiving grant reporting requirements for researchers who have to repeat experiments. The Texan institutions are adding up the damage for insurance firms, with the total expected to run into millions of dollars.

Emergency generators were still providing power at the Baylor facilities a week after the storm, with lab directors concerned that freezers holding specimens and cultures could overheat. Before the generators were brought in, students, postdocs and professors carried tonnes of dry ice up flights of stairs to keep the specimens frozen.

Congress cautious on advice office

Matthew Davis, Washington

Science-policy experts met in Washington on 14 June in a bid to revive the Office of Technology Assessment (OTA), the science and technology advisory unit that Congress disbanded in 1995.

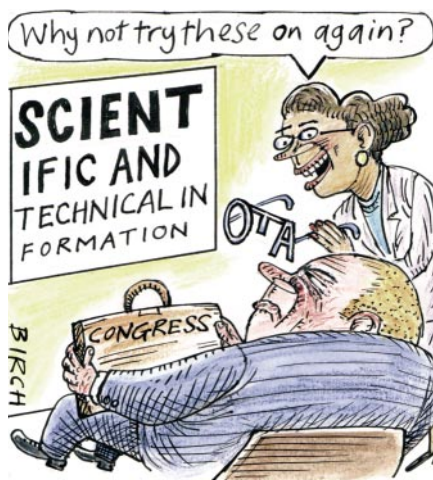
But before the meeting, which discussed five alternative models for the office, congressmen offered up some advice of their own: don't worry too much about the details, and don't expect action on the matter any time soon.

With law-makers facing a raft of difficult science-related issues, many academics and policy experts say that Congress should restore something akin to the OTA.

But the chair of the House Science Committee, Sherwood Boehlert (Republican, New York), said that Congress already has too much scientific information. He added that it can also turn to the National Academy of Sciences for detailed advice, and expand the role of the Congressional Research Service and the General Accounting Office if more input is needed.

Boehlert nonetheless co-sponsored a bill introduced last week by Rush Holt (Democrat, New Jersey) to revive the OTA and give it an annual budget of \$20 million.

But Vernon Ehlers (Republican, Michigan) warned that many Republicans



still think of the OTA as serving Democrats' partisan interests and producing reports too ponderous for practical use. He said that Congress instead needs a means of tapping technical expertise at two or three days' notice, given the speed of legislation.

Granger Morgan of the department of engineering and public policy at Carnegie Mellon University in Pittsburgh, who organized the meeting, says Congress needs an independent body to give it a sound basis for decision-making. "They need more than bare facts and brief interactions with technical experts," he says.

Japan aims to forge stronger European links

David Cyranoski, Tokyo

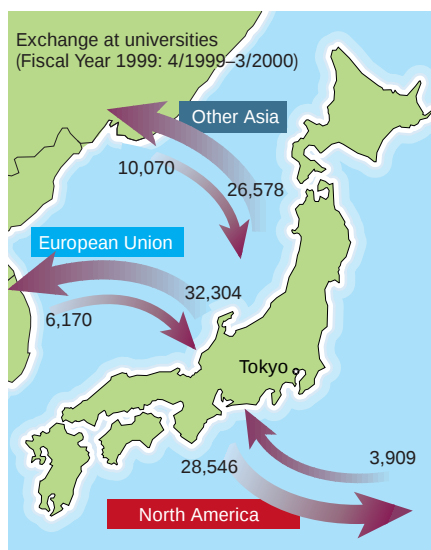
Government officials in Japan are addressing what they see as a growing problem in science — weak links with Europe and a chronic shortage of European researchers willing to visit Japan.

Although Japan sent more than 32,000 researchers to European universities in 1999, only 6,170 Europeans went to Japanese universities, according to government figures. An analysis of co-authored papers by Masamitsu Negishi of Japan's National Institute of Informatics shows that relations between Japan and the European Union (EU) are by far the weakest link in the US–EU–Japan triangle that dominates global research.

Demand from Europeans for fellowships to work in Japan has been moribund, a Tokyo workshop organized last month by the Japan Society for the Promotion of Science (JSPS) was told. The society offers 20 fellowships a year, with benefits including a tax-free stipend of ¥400,000 (\$3,250) a month. But in the past three years, only two researchers applied and only one took up the fellowship.

Explanations include low awareness of the fellowships, fear of the language barrier and concerns that Japan is weak in some areas. "There is a perception in Europe that going to Japan is very risky for your career, because Europeans underestimate Japan's research level," says Jan Carlstedt-Duke of the Karolinska Institute in Stockholm.

Japan's plans include the establishment of



Ebb and flow: movement of researchers in 1999.

a JSPS office somewhere in the EU, and the broadening of the '2 + 2 programme', which allows a researcher to spend two years in Japan with the assurance of two years of funding back home afterwards. The scheme is being piloted in the United Kingdom.

Collaborative institutes, such as the Laboratory for Integrated Micro-Mechatronics Systems, a joint project between France's research agency, the CNRS, and the University of Tokyo, may also be expanded.

But some European researchers blame

the lack of a 'collaborative research culture' in Japan. "It is difficult, because of language and other barriers, to have interaction with mentors or with other postdocs and graduate students," says one former postdoc.

Although Japan has some of the best-equipped laboratories in the world, lack of technicians can leave researchers out in the cold. "So you're sitting there with the best machine around and a manual, probably in Japanese. It's hard to move forward," says a European postdoc working on nanostructures at Japan's National Institute of Advanced Industrial Science and Technology, who nonetheless says the scholarship was "a unique opportunity to broaden my understanding of various fields".

Some participants recommended offering fewer but more remunerative scholarships. They said Europeans were deterred by the fact that (unlike Japanese researchers visiting Europe) they won't get their jobs back when they return. Very few visitors to Japan obtain tenured university positions there.

Officials from Japan and the EU will meet later this year to try to agree on concrete steps to address the problem.

Meanwhile, Japan has established a liaison office in Sweden in an attempt to boost Japan's international scientific standing in line with a government-set target of 30 Nobel prizes over the next 50 years.

► <http://www.fujita3.iis.u-tokyo.ac.jp/~limms>

► <http://www.jspis.go.jp/e-home.htm>

Stem-cell research delayed by German ethics council

Alison Abbott, Munich

Germany's main university granting agency has bowed to political pressure and agreed to delay its decision on whether to fund research using human embryonic stem cells.

The National Ethics Council, recently established by Chancellor Gerhard Schröder, asked the Deutsche Forschungsgemeinschaft (DFG) to postpone a decision, to give it time to formulate an opinion.

The DFG stated at the beginning of May that it was ready to fund the research. But in response to pressure from the federal research ministry, which called for a full public debate, it delayed until July a decision on its first application, from Oliver Brüstle and Otmar Wiestler at the University of Bonn (*Nature* 411, 119–120; 2001).

Following the council's request, however, the DFG said it recognized that public debate needed to continue, and agreed to put the decision on ice until at least October.

"From a scientific point of view, nothing has changed our mind about the

appropriateness of the DFG funding embryonic stem-cell research," says Claus Bartram, a human geneticist from the University of Heidelberg and a member of the DFG's committee for basic questions in gene research. "But the political pressure has been too great."

Bartram is frustrated by the delays, which he says could mean the shelving of the Bonn team's research plans. Even as the stem-cell debate floods the German media, Bartram says: "No new points of view are being put forward. All the arguments are already on the table. It is time for decisions to be taken."

Germans are particularly sensitive about using embryonic stem cells for research because of human-research abuses during the Nazi era. Many argue that such cells should be subject to constitutional human rights, and would-be researchers have been compared to Josef Mengele, who performed horrific genetic experiments at Auschwitz.

The issue has divided political opinion across party lines, and at the highest levels.

Schröder supports the research, whereas President Johannes Rau, who is also a Social Democrat, opposes it. Wolfgang Clements, the prime minister of Nordrhein–Westfalen, where Bonn is located, has offered to pay for cells to be imported from Israel for Brüstle and Wiestler to use.



Cell mates: Wiestler (left) and Brüstle have had their hands tied by the ethics debate.

Ruling makes it harder to convict dig thieves

Rex Dalton, San Diego

Archaeological sites across the United States are under increased threat from looters and vandals, according to researchers in the field. They argue that a ruling last December by a Californian court has made it harder to secure convictions under the federal Archeological Resources Protection Act (ARPA).

The ruling raised the burden of proof for prosecutors, requiring them to prove that a defendant knew that a looted specimen was an archaeological resource. Archaeologists fear that this decision will limit prosecutions under ARPA, which was enacted in 1979 to help save cultural artefacts situated on federal land.

As if to confirm these worries, a federal prosecutor in Colorado last month declined to use ARPA to indict a couple alleged to have removed Native American skeletal remains from a lake bed in the San Juan National Forest near Dolores, Colorado. The assistant US attorney in Durango, Robert Kennedy, blamed the court ruling as the reason for not proceeding with felony indictments. The case remains under investigation, but some federal authorities fear that there will be no prosecution. Kennedy declined to be interviewed.

Archaeologist Mark Varien, research director at the Crow Canyon Archaeology Center in Cortez, Colorado, described the failure to prosecute as “the height of absurdity”, as it is widely known in the region that it is illegal to take archaeological artefacts. “This gives a green light to those who want to loot,” says Varien, “and it fails to educate the public about damaging these dwindling resources.”

At the annual meeting of the Society for American Archaeology (SAA) in New Orleans this spring, researchers and law-enforcement officers held a symposium on ARPA's role. From 1906 until 1979, authorities say there were only 18 criminal convictions for archaeology crimes on federal lands under a previous law. But during the past 10 years, there have been at least 37 convictions under ARPA, including several in which commercial looters were imprisoned for years.

“The ARPA statute has not solved the archaeological-resources crime problem, but it has had a significant impact on the activities of looters and vandals,” says Martin McAllister, a former US Forest Service archaeologist who now is a Montana consultant on archaeology crimes.

But the case of Ian Lynch, who has been charged under ARPA for removing a 1,400-year-old skull from Heceta Island, Alaska, could undermine the law's track record. The US Court of Appeals for the Ninth Circuit



Hunting ground: McPhee Reservoir in Colorado is a favourite target of archaeological looters.

in San Francisco ruled last December that prosecutors must prove that Lynch knew the skull was an archaeological resource. It is this standard of proof that has upset archaeologists and law-enforcement personnel.

Wayne Dance, an assistant US attorney in Salt Lake City who has prosecuted some of the nation's most celebrated archaeological cases, called the Lynch decision “fundamentally flawed”. As a result of the decision, Dance says, prosecutors will have “to prove that the defendant knew that his or her conduct affected an archaeological resource, regardless of the otherwise

wrongful nature of that conduct”.

Privately, a federal prosecutor in another southwestern state said that the government may now need either a confession or direct evidence — such as a tape recording or writing — of a defendant's knowledge that the objects in question are an archaeological resource.

But Steven Skrocki, the Assistant US Attorney in Anchorage, Alaska, who is prosecuting the Lynch case, believes that he will still win a conviction of the defendant, who now is pleading innocent for a trial set for later this year. ■

Deputies caught on the wrong side of the law

Looters of archaeological sites in the United States are often petty criminals with a history of convictions for theft and other offences. But an Arizona case has exposed a different breed of ‘pot-hunters’: a couple of sheriff's deputies, one of whom learned archaeology from a government scientist.

The two Yavapai County deputies lost their jobs after pleading guilty last month to a felony charge of looting an archaeological resource.

Tony Mascher and John Price will be sentenced next month in the US District Court in Phoenix for violating the Archeological Resources Protection Act at the

Kinnikinick site in the Coconino National Forest. They face possible prison sentences.

As a high-school student, Mascher opted to study archaeology under a Forest Service archaeologist in the Tonto National Forest, northeast of Phoenix. Federal authorities believe he used this knowledge for his later criminal behaviour.

After searches of Mascher's home a year ago, authorities found an extensive collection of “museum-quality” Native American artefacts. All were seized by federal agents. However, most are expected to be returned to Mascher, as he has told authorities that they

were legally obtained from private land.

The sheriff's deputies were caught when an Arizona State University archaeology graduate student, Wesley Bernardini, discovered them digging at the isolated Kinnikinick site. Using their legal knowledge, the deputies delayed a search of their camp-site and vehicles until late at night. The authorities believe this allowed them to throw away incriminating artefacts. When federal agents finally examined the camp-site, they found it surrounded by discarded pieces of pottery, some of which matched freshly excavated artefacts from Kinnikinick. R.D.

Europe plans concessions to kick-start Kyoto

Jim Giles, London

In a bid to activate the Kyoto protocol on climate change without the United States, negotiators from the European Union (EU) are planning to offer concessions to Japan at next month's meeting of the parties to the treaty in Bonn.

The protocol cannot come into effect until it is ratified by countries that together account for 55% of greenhouse-gas emissions from developed nations, as measured in 1990. Europe, including non-EU countries and Russia, accounts for about half of these emissions, but it would probably need the cooperation of Japan, which accounts for 8.5% of them, to reach the threshold and bring the treaty into effect.

Japan has previously said that it will not ratify the accord unless the United States also ratifies it. But the Europeans hope to persuade it to do so on the basis that the United States will eventually rejoin the process.

The EU's concessions were released on 11 June by Jan Pronk, the Dutch environment minister, who will chair talks on the protocol when they resume in Bonn next month.

The document alters the amount by which Japan's emission target can be reduced if the country changes its land-use patterns. The protocol commits most developed nations to overall cuts of between 5% and 8% on 1990 emission levels by 2012. Countries can earn reductions in their target by a variety of measures, such as planting new forests, which absorb carbon dioxide.

The new text offers Japan an exemption from some of the restrictions controlling the effect of land-use changes on the emissions target. The move responds to Japan's claim that it is difficult for it to make cuts by other means. It has argued, for example, that it is hard for it to raise energy efficiency, because its cars and factories are already so energy efficient. Japan has little free land for new forests, but thinks that it can increase the amount of CO₂ absorbed by its existing forests.

But the concessions fall short of what Japan was seeking — according to one official, they would allow a 3% cut from land-use changes, whereas Japan wanted 3.7% — and it remains to be seen whether they will be enough to bring the country on board.

Japanese officials say that the country has not yet given up hope that the United States will ratify the protocol, and that it will not decide what to do until after talks between Prime Minister Junichiro Koizumi and US President George W. Bush later this month.

The future of the protocol has been in doubt since Bush declared in March that the United States would not ratify the treaty. But EU heads of state, after meeting with Bush in Gothenburg, Sweden, last week, confirmed that they will press ahead with ratification.

EU negotiators say that they remain confident that the Bonn talks will move the protocol forward, even without US involvement. "I would like to see the protocol ratified at the 2002 World Summit on Sustainable Development in Johannesburg,"



Out in the cold: protesters outside the European summit signal their anger at President Bush.

says Yvo de Boer, a member of the team of European officials that drafted the changes.

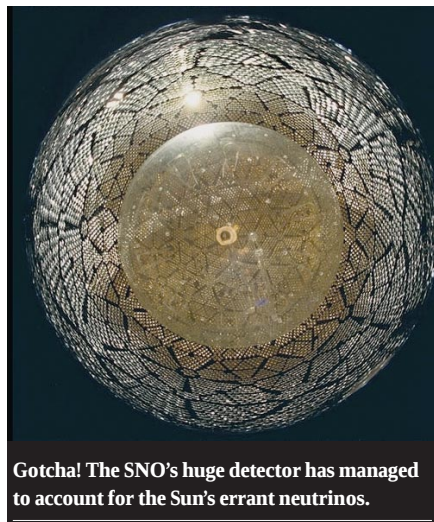
EU officials say that they hope the United States will sign up to a second agreement to cover emissions in the period after 2012. Talks on this accord should begin within the next two years.

Flavour switching solves riddle of missing neutrinos

The solar neutrino problem has been solved, according to researchers working at the Sudbury Neutrino Observatory (SNO) in Canada. Their results explain why physicists looking for neutrinos emitted by the Sun have detected between a half and a third as many particles as predicted.

The Sun mainly emits electron neutrinos, but the particles also come in two other 'flavours', known as tau and muon neutrinos. Neutrino detectors are primarily sensitive to electron neutrinos, which means the deficit could be explained if some electron neutrinos were changing into tau or muon neutrinos en route to the Earth.

SNO, a joint venture between Canada, the United States and Britain, is a spherical vat containing 1,000 tonnes of heavy water, located two kilometres below ground in a mine some 300 kilometres north of Toronto (see *Nature* 411, 10–12; 2001). As currently configured, it can accurately measure the flux of electron neutrinos reaching the



Gotcha! The SNO's huge detector has managed to account for the Sun's errant neutrinos.

Earth. So by comparing these results with those from other detectors — which have a small sensitivity to tau and muon neutrinos — the SNO team has shown in a paper

submitted to *Physical Review Letters* that some electron neutrinos emitted by the Sun are switching flavours.

"It's an enormous achievement," says John Bahcall of the Institute for Advanced Study in Princeton, New Jersey, who works on the solar neutrino problem. "I'm thrilled."

The result has important consequences for the standard model, the theory that describes fundamental particles and their interactions. Theorists have assumed that neutrinos have zero mass — but neutrinos must have mass if they are to switch flavours. "These results go beyond the standard model," says Art McDonald, director of SNO. "Now they have to be explained."

SNO will now be reconfigured to be sensitive to all three classes of neutrinos, providing further insights into flavour switching.

► http://owl.phy.queensu.ca/sno/first_results

*A News and Views article on the SNO results will appear in a future issue.

Canada opens bidding to host Iter fusion research facility

Moscow Canada has become the first country to officially offer to host the Iter fusion energy research facility, formerly known as the International Thermonuclear Experimental Reactor.

Iter is a joint international effort to test the feasibility of fusion power. The United States withdrew from the project three years ago and some fusion researchers think that building the facility in Canada could tempt the United States to rejoin.

Canadian officials made their offer at a meeting of the consortium held in Moscow to mark the completion of Iter's engineering design phase. Started in 1988, this phase cost US\$1 billion to complete and produced a 10,000-page design report. Iter would cost an estimated US\$3.5 billion to build.

The Canadian bid may stimulate rival offers from the other members of the consortium. French fusion researchers are keen to host the reactor at a nuclear research centre at Cadarache near Aix-en-Provence, and the Japanese government is currently considering three possible sites.

Canada's bid is strengthened by the fact that it can offer the world's only major

source of tritium, which is needed to fuel Iter and is already extracted from heavy water at a power station next to the proposed site at Clarington, near Toronto. But many fusion scientists are sceptical about the bid, because Canada, unlike Japan and the European Union, has no large-scale fusion research programme from which it could channel resources to pay for the project.

♦ <http://itercanada.com>

Transgenic corn cleared of link to human allergies

Washington There is no evidence that genetically modified StarLink corn caused allergic reactions when it entered human food products, a report from the US Centers for Disease Control and Prevention (CDC) said last week.

StarLink corn, which is engineered to produce an insecticidal protein, is licensed for use only in animal feed because of fears that the protein, known as Cry9C, could cause allergic reactions in humans. But StarLink was found in food products such as taco shells last year (see *Nature* 407, 438; 2000).

The US Food and Drug Administration subsequently received reports of several potential allergic reactions to foods that may have contained StarLink. The CDC then took



No reaction: StarLink corn gets the all-clear from the US government over allergy accusations.

blood samples from 17 people who had complained of reactions, but failed to find antibodies against Cry9C in any of them.

♦ <http://www.cdc.gov>

French council backs human cloning ban

Paris France seems set to maintain its ban on the cloning of human embryos after the Conseil d'Etat, one of three bodies that makes up the French supreme court, rejected a government proposal to make it legal for therapeutic purposes.

The decision is part of a review of French bioethics legislation. Prime Minister Lionel Jospin surprised politicians and scientists last November when he backed an updated bill that permitted the cloning of human embryos for the creation of embryonic stem cells (see *Nature* **408**, 629; 2000). The bill is yet to go through parliament, but was rejected by the national human-rights commission and only narrowly approved by the national bioethics committee.

Jospin is now expected to follow the opinion of the Conseil d'Etat in order to avoid a public debate on the issue in the run-up to next year's presidential elections. His likely opponent, current President Jacques Chirac, has already declared his opposition to therapeutic cloning.

Doctoral students revolt over ten-year pay freeze

Paris A government refusal to end a ten-year-old pay freeze brought doctoral students onto the streets of Paris on 7 June. Their monthly stipend remains fixed at FF7,400 (US\$970).

Around 700 doctoral researchers protested in front of the Ministry of Economy, Finances and Industry in Paris. A call for a boycott of laboratories has been made for the 28 June.

"Doctoral students are being put in

competition with researchers because the Ministry of Research says that any increase in stipends can only come from a decrease in laboratory budgets," says Claire Poinot, president of the Confederation of Student Researchers.

New potato gives hope in Russian blight fight

Munich Russia's battle against potato blight has been boosted by the development of a potato variety that is resistant to potato late blight, scientists said at an international workshop in Warsaw, Poland, on 6–9 June. Nutrition experts have warned that Russia faces a potentially catastrophic potato famine (see *Nature* **410**, 1011; 2001).

The new variety, known as New York 121, is to enter field trials in the St Petersburg and Moscow regions. If successful, seed potatoes could be sold to millions of small farmers in Russia. Potato blight currently destroys more than 10% of the country's crop each year.

NASA shortlists missions to Mars and Pluto

Washington A dozen proposed spacecraft missions — ten to Mars and two to Pluto — have been selected by NASA for further study and a chance of a launch later in the decade.

The ten Mars projects include a camera-equipped glider, a network of weather stations on the martian surface, and a satellite that would trap atmospheric dust with a flypaper-like aerogel and return it to Earth. The first choice is due to fly in 2007.

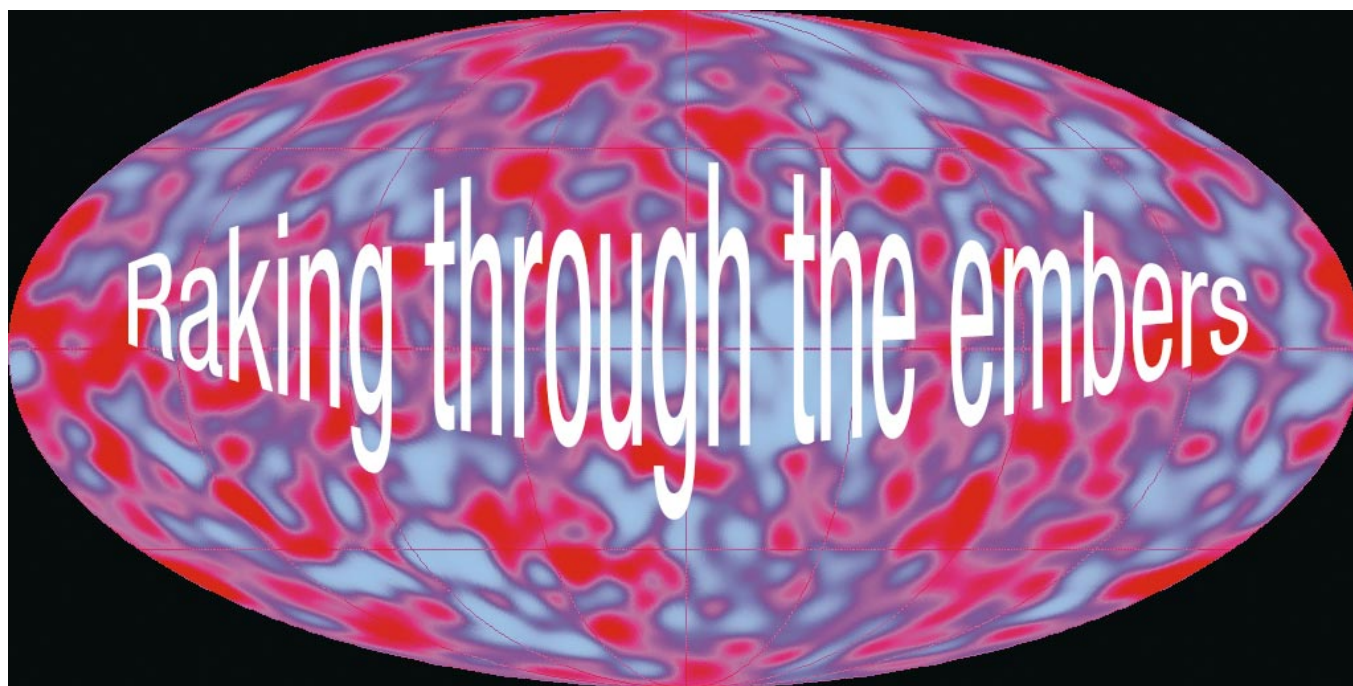
The teams behind the two Pluto missions have three months to refine their concepts. By then NASA should know whether Congress has plans to fund a Pluto mission, which the Bush administration omitted from next year's NASA budget request.

Universities go head-to-head over donation

Boston Harvard and Stanford universities are competing for a \$150-million donation from Larry Ellison, chief executive of Oracle Corporation in Silicon Valley, California. The money will be used to establish an institute to study the impact of technology on society.

The success of Oracle, a computer-database company, has made Ellison one of the world's richest men. He says he chose Harvard and Stanford because of their strengths in technology, economics and politics.

No firm timetable has been set for a decision, although it could be reached within a few months, says Jennifer Glass, a spokeswoman for Oracle.



NASA/COBE

Clues about the origin and fate of the Universe lie hidden in the microwave radiation left over from its early days. Tom Clarke examines the latest attempts to map the Big Bang's afterglow.

The first high-resolution images of the radiation that makes up the cosmic microwave background (CMB) earned rave reviews. In 1990, the American Astronomical Society gave an early presentation on the new images a standing ovation. Two years later, Stephen Hawking of the University of Cambridge described one result¹ arising from the image data as the "discovery of the century".

Hawking was praising an analysis of a CMB image taken by NASA's Cosmic Background Explorer (COBE) satellite. The CMB provides a direct link back to the early Universe — it is a cosmic 'baby picture' that faithfully reproduces every freckle and dimple of the infant Universe. COBE's data showed that not long after the Big Bang, matter was starting to clump together. Cosmologists believe that this uneven distribution of matter was the starting point for the formation of stars and galaxies.

But these results were just the beginning for CMB research. Thanks to plans for two successors to COBE — one of which is scheduled to launch next week — new, higher-resolution images will soon be available. Armed with these, cosmologists will be able to test their understanding of what happened after the Big Bang and, perhaps, explain why the Universe looks the way it does today. "We're entering the decade of CMB science," says Charles Lawrence, a CMB physicist at the Jet Propulsion Laboratory in Pasadena, California.

The CMB dates back to a time when the Universe was just 300,000 years old. Back then, the Universe was just a soup of photons, electrons, protons and helium nuclei — atoms were yet to form. The electrons were locked in a frantic tango with the photons, constantly absorbing and re-emitting them.

But as the Universe expanded and cooled, protons and electrons began to pair up to form hydrogen atoms. This freed the photons, as electrons in hydrogen atoms are much more reluctant to interact with photons. Physicists describe this as the 'decoupling' of photons and electrons. The decoupled photons exist to this day as the CMB, a blanket of microwave radiation that pervades every corner of the sky. "It's our cleanest probe of the early Universe," says Max Tegmark, a cosmologist at the University of Pennsylvania in Philadelphia.

Treasure maps

The CMB contains a wealth of information about the state of the Universe at the moment of decoupling. COBE's big contribution was to reveal differences of around 1 part in 10,000 in the average energy, or temperature, of CMB photons found in different areas of the sky. COBE's maps (see main picture) of the CMB's temperature show ripples of alternating hot and cold areas. These arise because at the moment of decoupling different parts of the Universe had different densities — the hotter CMB photons came from the denser areas of the early Universe. The

presence of these ripples added support to the idea that stars and galaxies formed from an uneven distribution of matter.

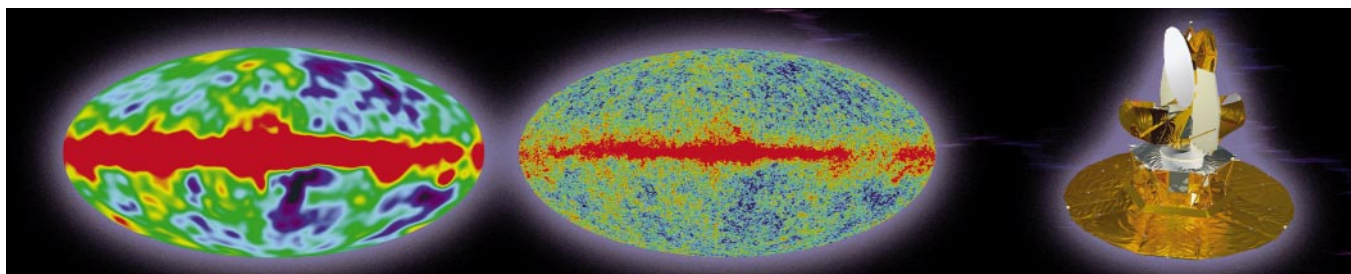
But COBE's sensors lacked the resolution needed to answer many important questions. A finer-grained analysis of the CMB will provide cosmologists with the missing details, and should help them decide which of the proliferating models describing the evolution of the Universe is correct.

The significance of the information tied up in the CMB is such that ground-based projects have rushed to "skim off the cream" of CMB data before the launch of the space missions, according to Lyman Page, a



Twin peaks: Lyman Page feels that two missions investigating the CMB are better than one.

ANGEL OTAROLA



Blowing hot and cold: with the launch of NASA's MAP satellite (right) next week, cosmologists hope that the image data of temperature fluctuations in the cosmic microwave background produced by the COBE satellite (left) will be replaced by a much higher-resolution version (simulated, centre).

physicist at Princeton University in New Jersey. Unlike COBE, sensors on ground stations and balloons can only map a restricted area of the sky, but that has not stopped them from producing some important results.

Last year, the international Balloon Observations of Millimetric Extragalactic Radiation and Geophysics (BOOMERANG) experiment — a sensitive microwave detector slung beneath a balloon which circled the Antarctic — achieved the best picture yet of the CMB². Analysis of these data confirmed that the Universe is 'flat' — it will continue to expand for ever, rather than collapsing in a 'big crunch' as some models had predicted.

Vive la resolution

But now the main focus of activity is poised to shift from ground- and balloon-based experiments to two satellites that promise to study the CMB with unprecedented accuracy. Space-based experiments have some big advantages. The new probes will take CMB research away from the noisy microwave sources on Earth, such as cellphones and radar systems, as well as avoiding the scattering effects of the atmosphere. And, like their forebear COBE, they will scan the entire sky, but this time at much higher resolutions.

Ripples with an angular resolution of less than 7° were invisible to COBE (1° is about twice the diameter of the Moon). BOOMERANG's limited sky view had a resolution of around 0.25° . But if all goes well, next week will see the launch of NASA's Microwave Anisotropy Probe (MAP), which can conduct a full-sky survey of variations down to a scale of 0.3° . And in 2007, the European Space Agency (ESA) plans to launch Planck, an even more sophisticated CMB probe which the agency hopes will map the CMB at a resolution of 0.17° .

The images that MAP and Planck will produce are eagerly awaited by cosmologists. According to the theory of how the variations in the CMB formed, ripples at different scales are linked to different fundamental properties of the early Universe.

Ripples at around the 1° scale have their origins in the oscillations that passed through the early Universe. Gravitational forces tried to pull the soup of matter inwards, but this was resisted by pressure arising from the movement of photons. The push and pull of

these forces produced a series of reverberations and created patches of high and low density — the cause of the small-scale ripples.

Analysis of these ripples will tell cosmologists about the structure of the soup of matter through which they once moved. Researchers hope that data from MAP and Planck will give them better estimates of properties of the early Universe, such as the density of protons and electrons, and the distribution of energy between matter and radiation.

These more accurate figures should help to eliminate some of the competing descriptions of how the Universe evolved. Most cosmologists are convinced that the Universe underwent a period of 'inflation', during which it was expanding faster than the speed of light, before decoupling. But measurements of many of the parameters in the competing inflationary theories have large errors associated with them, making it difficult to determine which one is correct. "You really have to get very precise if you want to back the theorists into a corner," says Philip Mauskopf, a member of the MAP team who is based at Cardiff University in Wales.

Double vision

Despite the importance of CMB data, the high cost of space missions does bring into question the need for two spacecraft. The probes were approved by their respective agencies within a month of each other in 1996. Some researchers suggest privately that the two agencies, having made CMB science their top astrophysical priority, lacked confidence in one other's selection procedures. Scientists affiliated with the MAP team say that they never believed that Planck — being expensive and containing a number of untested technologies — would survive the ESA's approval process.

Whatever the reasons, CMB scientists are happy with an outcome that gives them two spacecraft, allowing the teams to double-check their data. "The CMB is so important that two missions would be necessary anyway," says Page, who is a member of the BOOMERANG and MAP teams.

There is also a small possibility that new maps of the CMB will rule out the inflation theories altogether. The initial data from Earth-bound experiments seem to be in agreement with inflation, but there is still too

much noise in the signal to rule out an error.

But a crucial test of the theories may need more data than MAP or Planck can provide. Inflation makes specific predictions about the polarization of CMB photons, which could be used to rule out rival theories. Planck will measure polarization, but only down to a scale of 10° . More accurate measurements will be made by two balloon-based experiments — extensions of the BOOMERANG and international Millimeter Anisotropy Experiment Imaging Array (MAXIMA) projects — but only over restricted areas of the sky. Together, the satellite- and balloon-based polarization surveys will be an important step towards testing inflation — but a high-resolution, full-sky survey will have to wait for a future, as yet unplanned mission.

Most researchers believe that inflation will survive the decade of CMB science. But they agree that the next ten years should see a cull of competing cosmological theories as researchers move towards a consensus on how the Universe evolved.

Tom Clarke works in Nature's science writing team.

1. Smoot, G. et al. *Astrophys. J. Lett.* **396**, L1–L4 (1992).
2. de Bernardis, P. et al. *Nature* **404**, 955–959 (2000).

Web links

Microwave Anisotropy Probe

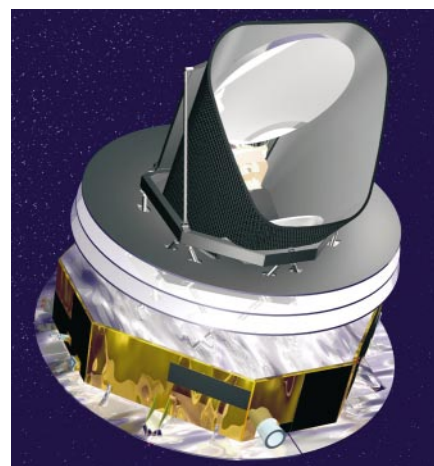
► <http://map.gsfc.nasa.gov>

Planck

► <http://astro.estec.esa.nl/SA-general/Projects/Planck>

BOOMERANG

► <http://www.physics.ucsb.edu/~boomerang>



Mapping the past: set to launch in 2007, Planck will offer the highest-resolution CMB images.

The search for autism's roots

A marked rise in the number of people diagnosed with autism sparked controversy over the safety of certain vaccines. As the furore dies down, Kathleen Wong talks to researchers seeking the real causes of this unsettling condition.

In child-development circles, it is known as the new epidemic. Anxious parents are flooding clinics with children who won't speak, who scream at the touch of clothing on their skin and who seem locked in a world of their own. The diagnosis is autism.

In the mid-1970s, autism was seen as a rare condition, thought to affect between two and four children in 10,000. But recent studies suggest that the figure has grown, on average, by 40% (ref. 1). Most autism researchers think that changes in the way the condition is defined and diagnosed lie behind the rise in recorded cases. But that is of little consolation to the parents of children with autism. They want to know what is happening to the minds of their children and whether anything can be done to treat the condition.

Hopes that these questions will be answered are now higher than ever. New technologies are offering deeper insights into the way the autistic brain functions, and advances in molecular biology promise to unearth the disorder's genetic roots. Meanwhile, educators and therapists are discovering that, if caught early, children who would have been institutionalized just 30 years ago can now stay with their families and attend mainstream schools.

Unusual behaviour

Autism can seem a somewhat paradoxical disorder. Around three-quarters of all people with autism are mentally retarded², but a small fraction have considerable mathematical or artistic abilities. Some are insensitive to pain and have a tendency to injure themselves; others are hypersensitive to touch. People with mild forms of autism can have strong verbal skills, whereas those with more severe forms are unable to speak at all.

Underlying these behaviours are problems with social interaction and communication and a tendency to take an obsessive



Under analysis: studies of brain activity are revealing the ways in which the minds of children with autism process information (top), and therapy is helping to improve the children's social skills.

interest in particular subjects. The first two features of autism can appear at a very young age. Babies with autism seem to lack the innate preference that normal babies have for looking at faces rather than other objects³. They fail to develop the skills needed to understand the subtleties of facial expressions, a vital part of face-to-face communication. As they grow older, children with autism act as if they are cut off from the world. Researchers using brain-imaging techniques have found that these children process visual information about faces in the same part of the brain they use to process information on objects⁴ — most people use other, specialized, brain regions to perceive faces.

In addition, children with autism seem unable to guess what other people might know or are thinking by studying what they say and do. 'Sally-Anne' scenarios used by psychologists highlight this deficit. In one example, Sally and Anne are playing together. Sally places a marble in a covered basket and

leaves the room. Anne then moves the marble from the basket to a box. Most children with autism think that when Sally returns, she will look for the marble in the box⁵ because they cannot view the situation from Sally's perspective. Normal children can anticipate Sally's behaviour because they have a 'theory of mind', which usually develops at around age four⁶. But many children with autism never develop such an understanding.

Individuals with autism also have a tendency to focus on details, rather than the big picture. Such children might concentrate on the colour of a utensil rather than its shape, preventing them from learning the difference between a fork and a spoon. This love of detail can, in a few rare cases, result in extraordinary talents — Stephen Wiltshire, a British artist with autism, is famed for his ability to produce beautifully detailed sketches of complex buildings.

Recognition that a range of symptoms is associated with autism has forced psychia-



ducted in Japan, the United States, Britain and Scandinavia⁷⁻¹⁰ show a rise in the number of cases over the past 45 years.

But take into account changes in the way the condition is defined, and the figures do not seem so disturbing. The first epidemiological studies, conducted in the 1960s and 1970s, used narrow definitions of autism that excluded individuals with milder forms of the disorder. The earliest definitions also omitted mentally retarded children who may also have had autism. "It would be very surprising indeed if the broadening of the criteria for autism weren't the major part of the explanation," says Michael Rutter of the Institute of Psychiatry in London.

Increased public recognition of the disorder is also likely to have contributed to the apparent epidemic. As parents and doctors have become more familiar with the disease, the chances that they will identify potential cases and refer them to psychiatrists have increased. The rising number of children attending nursery schools, where they get more attention than at other stages of their schooling, may also play a role, as a child's social deficits tend to stand out from the behaviour of its peers.

Increased awareness

The growing consensus that changes in diagnosis and public awareness are behind the rise in recorded cases came as a relief to public health officials responsible for vaccination programmes. In 1998, Andrew Wakefield, a gastroenterologist at University College London, proposed a novel and terrifying connection between autism and the combination measles-mumps-rubella (MMR) vaccine¹¹. He described the cases of 12 children who appeared to be developing normally until they received their MMR shot between the ages of 15 and 18 months. Soon after, the children developed a kind of inflammatory bowel disease, began losing basic speech and social skills, and were subsequently diagnosed with autism.

The theory found a ready audience among parents and the press. Around half of all children later diagnosed with autism appear to develop the symptoms very suddenly, around the time that the MMR jab is



Not guilty: the idea that the MMR vaccine might give rise to autism has now been rejected.

administered. These children regress severely over a period of six months, losing social skills that they had begun to master. The fact that the vaccine first went into widespread use in the 1980s, when studies suggest that incidence rates for autism started to climb, added credibility to the claim.

Subsequent studies have failed to find a link between MMR and autism¹²⁻¹⁴. On closer analysis, the data from several parts of the world show that the rise in autism actually started before MMR, Rutter explains.

But just as the MMR controversy is dying down, another potential vaccine-related cause for autism has been highlighted. Many vaccines use a mercury-containing preservative called thiomersal. Fears that vaccinations may be exposing children to dangerous levels of mercury have led the US Institute of Medicine to schedule a meeting for next month to discuss possible links between thiomersal and autism. Most researchers doubt that thiomersal will turn out to play an important role in the autism story, but they want to settle the issue to head off another vaccine scare.

Despite lingering fears over the safety of vaccines, many researchers believe that the real key to understanding autism lies in sufferers' genes. Studies of twins offer the most convincing evidence of such a link. One study found that if one identical twin has autism, the other twin has a 60% chance of developing the condition, and a 92% chance of having a condition within the DSM's spectrum of related disorders. In non-identical twins the odds of the second twin having an autism-spectrum disorder fall to 10%. The study failed to find any cases in which two non-identical twins both had autism¹⁵.

Studies of the inheritance of autism and the wide range of symptoms associated with the condition indicate that between three and ten genes are likely to be involved¹⁶. Some researchers suggest that mutations in

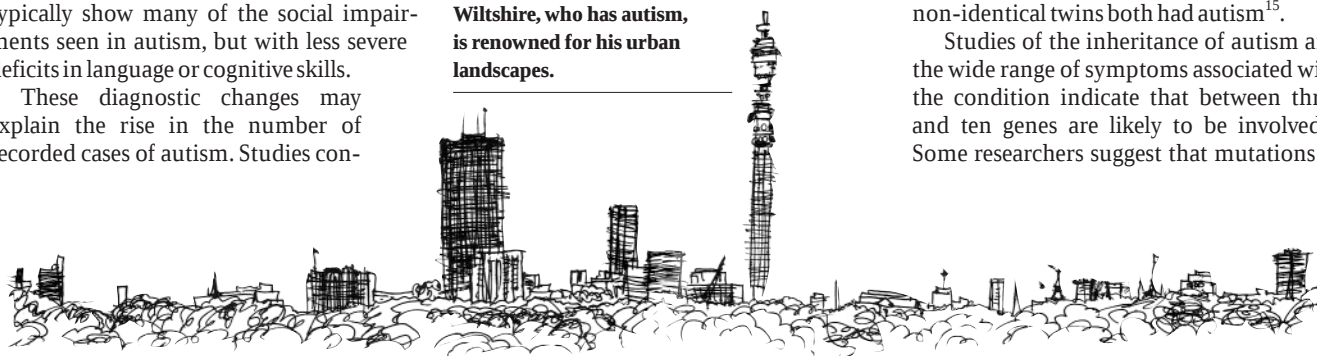


Simon Baron-Cohen suggests that autism may be an exaggeration of gender differences.

trists to change the way that they define the condition. Classical autism, first defined in the 1950s, is now seen as part of a spectrum of related conditions. The latest edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM)*, published by the American Psychiatric Association, recognizes five additional disorders, such as Asperger's syndrome, as part of the autistic spectrum of behaviour. People with Asperger's syndrome typically show many of the social impairments seen in autism, but with less severe deficits in language or cognitive skills.

These diagnostic changes may explain the rise in the number of recorded cases of autism. Studies con-

Eye for detail: artist Stephen Wiltshire, who has autism, is renowned for his urban landscapes.



more than a threshold number of these genes could cause classical autism, whereas fewer mutations would result in social shyness or delays in acquiring language skills.

The search for 'autism genes' has yielded some promising leads. Links have already been established with abnormalities on chromosomes 7 and 15 (refs 17–20). Many researchers are confident that at least one gene for autism will be identified within the next five years.

But the fact that the identical twin of a child with autism may not develop the condition suggests that environmental factors are also involved. Mutations to one or more autism genes may, for instance, increase a child's vulnerability to an unknown environmental trigger encountered during early infancy or in the womb.

Other researchers are searching for physiological signs associated with the disease, some of which seem to develop before birth. Autopsies of people with autism have revealed abnormalities that would have developed by around 30 weeks after conception²¹. In particular, structures in the brain's limbic system, responsible for emotions, aggression, sensory input and learning, appear to be underdeveloped.

A more low-tech analysis is also shedding light on developmental factors that may be linked to autism. Children with autism tend to have ring fingers as long as or longer than their index fingers²², as do their family members — an attribute that has been linked to high testosterone levels in the womb²³.

This finding also bolsters the 'extreme male-brain' theory of autism²⁴ put forward by psychologist Simon Baron-Cohen and his colleagues at the University of Cambridge. Baron-Cohen argues that many symptoms associated with autism are magnified versions of the normal gender differences between men and women. Men are known, for example, to be better at spatial tasks than women. Individuals with autism and those with Asperger's syndrome score even higher on these tasks than normal males. A similar trend is seen across a range of behavioural and biological attributes known to be linked to gender, from the size of particular areas of the brain to the age at which children acquire language skills. Baron-Cohen says that the theory is attracting attention from within the autism research community, but that it is too new to have been fully evaluated.

Other groups are looking for correlations between specific types of autism and mutations in potential autism genes. Gerard Schellenberg, a neurologist at the University of Washington in Seattle, and Geraldine Dawson, director of the university's Autism Research Center, are studying families in which at least two siblings have autism. The researchers aim to diagnose the specific types of autism the siblings suffer from and to link the symptoms to anomalies in their DNA.

Despite autism's complexity, researchers are confident that they will eventually understand it.

New therapies may eventually be developed from these studies. In the meantime, therapists are using what they already know about how the minds of children with autism function to design behavioural treatments.

Improved interaction

The most successful therapies focus on helping children with autism overcome their behavioural and communication problems. One example is the 'pivotal skills' programme developed by education researcher Robert Koegel of the University of California, Santa Barbara. The programme teaches children to ask questions instead of screaming for attention, and to respond to the salient features of an object, such as number and suit on a playing card, rather than the irrelevant details, such as a crease in the card's corner. Koegel motivates the children by choosing lessons suited to their interests, for example by teaching an animal-obsessed child to use terms for the difference between big and little by going to the zoo and comparing a rabbit and an elephant.

A lack of controlled studies makes it difficult to compare Koegel's programme with other behavioural approaches, but anecdotal evidence suggests that the system works well. "Even if they don't recover completely, they are overcoming the biggest problems," says Koegel. He now hopes to begin studying the neurological changes that take place in the

children's brains as they learn, and to match different therapies to different types of autistic disorders.

Early-intervention programmes such as Koegel's, along with the research into autism's causes, have raised hopes that autism may one day become a treatable condition. Despite its complexity, researchers are confident that they will eventually understand the disorder. Teasing out the genetic and environmental causes is a tricky task, but the many strands of research on this most baffling condition could come together over the next decade.

Kathleen Wong is senior editor of *California Wild*, the magazine of the California Academy of Sciences.

1. Fombonne, E. *Psychol. Med.* **29**, 769–786 (1999).
2. *Diagnostic and Statistical Manual of Mental Disorders* Fourth edn (Am. Psychiatric Assoc., Washington DC, 1994).
3. Joseph, R. M. & Tager-Flusberg, H. J. *J. Autism Dev. Disord.* **27**, 385–396 (1997).
4. Schultz, R. T. et al. *J. Arch. Gen. Psychiatry* **57**, 331–340 (2000).
5. Baron-Cohen, S. J. *Child Psychol. & Psychiatry* **30**, 285–297 (1989).
6. Wimmer, H. & Perner, J. *Cognition* **13**, 103–128 (1983).
7. Honda, H., Shimizu, Y., Misumi, K., Niimi, M. & Ohashi, Y. *Br. J. Psychiatry* **169**, 228–235 (1996).
8. Webb, E. V. J., Lubos, S., Hervas, A., Scourfield, J. & Fraser, W. I. *Dev. Med. Child Neurol.* **39**, 150–152 (1997).
9. *Prevalence of Autism in Brick Township, New Jersey, 1998: Community Report* (United States Centers for Disease Control and Prevention, 2000).
10. Gillberg, C., Steffenburg, S. & Schaumann, H. *Br. J. Psychiatry* **158**, 403–409 (1991).
11. Wakefield, A. J. et al. *Lancet* **351**, 637–641 (1998).
12. Taylor, B. et al. *Lancet* **353**, 2026–2029 (1999).
13. Kaye, J. A. *Br. Med. J.* **322**, 460–463 (2001).
14. Patja, A. et al. *Ped. Infect. Dis. J.* **19**, 1127–1134 (2000).
15. Bailey, A. et al. *Psychol. Med.* **25**, 63–77 (1995).
16. Rutter, M. J. *Abnormal Child Psychol.* **28**, 3–14 (2000).
17. International Molecular Genetic Study of Autism Consortium *Hum. Mol. Genet.* **7**, 571–578 (1998).
18. Cook, E. H. et al. *Am. J. Hum. Genet.* **62**, 1077–1083 (1998).
19. Cook, E. H. et al. *Am. J. Hum. Genet.* **60**, 928–934 (1997).
20. Schroer, R. J. et al. *Am. J. Med. Genet.* **76**, 327–336 (1998).
21. Kemper, T. L. & Bauman, M. L. *Neurol. Clin.* **11**, 175–187 (1993).
22. Manning, J. T., Baron-Cohen, S., Wheelwright, S. & Sanders, G. *Dev. Med. Child Neurol.* **43**, 160–164 (2001).
23. Manning, J. T., Scutt, D., Wilson, J. & Lewis-Jones, D. I. *Hum. Reprod.* **13**, 3000–3004 (1998).
24. Baron-Cohen, S. in *Neurodevelopmental Disorders* (ed. Tager-Flusberg, H.) 401–429 (MIT Press, 1999).



Pen pals: with therapy, children with autism are able to participate in mainstream schooling.

Poetry tells us that our souls have a shadow: can science respond?

Sir — Maurice Riordan's enjoyable, perceptive Words essay "The suspense of strangeness" (*Nature* **409**, 457; 2001) provides much information and insight. I would like to add some comments on how scientific training, method and language can be valuable to a poet. The compacted prose of scientific writing is training for writing concise poetry — especially in formats such as haiku. The necessary sitting, thinking and composition required to write poems can be made easier by the strict disciplines of scientific training.

Riordan writes from a poetic standpoint rather than, as I do, from a poet-scientist perspective. The perception that poets — and the general public — have of scientists is different from creative scientists' own perception of themselves. My definition of a poet is a person who has poems published after critical review by an independent publisher. Unfortunately, this produces the question: what is a poem? The Bristol Polygon poets in the United Kingdom would reply: "A poem is a poem when the person producing it says it is."

Some years ago, it was common in asthma conferences to schedule an afternoon to attempt to define asthma. The many discussions are crystallized by the comment "Asthma is like love: everyone knows what it is but no two people can agree the terms for its definition." Scientists know that the problem of definition is not confined to poetry or literature.

The answer to Riordan's provocative question "Why doesn't a scientific sense of wonder translate readily into poetry?" is that it can be and has been translated, but suffers the defects of all translocations and the inherent difficulty in writing good poetry. As a scientist, I appreciate that scientific discoveries and laws provide an imperfect insight into phenomena comparable to that partial vision provided by poetry. While scientists attempt to reveal or unveil the natural philosophy of the Universe, we should all be aware of how ineffectually we individually achieve that grand aim. Poets are often similarly ineffectual. Both groups have a literature which manifests their collective effectiveness.

Scientific poets may be more common than Riordan suggests: Alex Comfort, for example, published reputable poetry. And there's no bimodal separation into poets and non-poets: Siegfried Sassoon wrote that the night before his platoon went into action, during the First World War, a third

of the men became poets. In science we are familiar with the transformation of indifferent starters to excellent PhD graduates. I know of no evidence that the ratio of scientist-poets to scientists is any different from that of non-scientific poets to the non-scientific population.

However, creative scientists rarely have formal literary training. Accordingly, they are less likely than their pure poetic counterparts to be knowledgeable about what other poets have written. The standard of English language in many scientific publications is not high, offering little encouragement or example. Society's image of poets, or indeed of scientists, could make a scientist wary of being known as a poet for fear of reducing his or her scientific credibility. I believe that scientist-poets emerge from the woodwork with maturity or when they know their reputation is established.

Why should a reputable scientist bother about poetry? Gregory Corso (in *Penguin Modern Poets* volume 5, Cox & Wyman, London, 1963) explains:

*"I love poetry because it makes me love and presents me life ...
But it does tell me my soul has a shadow."*

Scientists can be and are poets. They may provide language and perspectives that are interestingly different from the general populace of poets.

Fleming Carswell

*Department of Pathology & Microbiology,
School of Medical Sciences, University Walk,
University of Bristol, UK*

Call to work together on microarray data analysis

Sir — Your Opinion article about gene expression, "Free and public expression" (*Nature* **410**, 851; 2001), and the accompanying News Feature about DNA microarrays, "When the chips are down" (*Nature* **410**, 860–861; 2001), reveal the tip of the iceberg.

Even larger problems loom ahead than the microarray data acquisition discussed in your articles. There is as yet no analytical standard to interpret the data. Traditional statistical algorithms are not sufficient for modelling functional genomics, and additional analytical problems abound, for example the issue of decoding the data structures that relate gene expression to gene function in complex biological systems.

We are undertaking a community-wide experiment to review state-of-the-art algorithms in microarray data analysis, to identify key research problems to be addressed. Last October, the Duke University Bioinformatics Shared

Resource initiated a critical assessment of microarray data-analysis techniques (CAMDA) which invited the whole scientific community to analyse the same standard data sets.

CAMDA is a functional-genomics successor to other community-wide experiments, such as GASP (<http://www.fruitfly.org/GASP1>) in genomics and CASP (<http://predictioncenter.llnl.gov>) in protein modelling.

CAMDA 2000 mobilized researchers in traditionally separate fields of study, creating an interactive forum for a review of microarray-analysis techniques. For the first time, molecular biologists, statisticians, mathematicians, computer scientists, engineers and physicists were all involved in looking at a problem whose ultimate solution will require their collaboration. The plan is to have a meeting every year to evaluate and stimulate academic research and commercial developments in microarray data analysis (see <http://www.bioinformatics.duke.edu/camda>).

Kimberly Johnson, Simon Lin

*Duke Bioinformatics Shared Resource, Box 3958,
Duke University Medical Center, 248 Hanes,
Durham, North Carolina 27710, USA*

Medical luminaries

Sir — May I add the names of more intellectual luminaries, whose careers were centred on Breslau/Wrocław, to the list offered by Min-Liang Wong ("Bright light of learning snuffed out in Breslau", *Nature* **410**, 865; 2001).

Ferdinand Julius Cohn (1828–1898), the prominent botanist and micro-biologist, father of bacterial taxonomy, was born in Breslau and became a professor at the university there. Julius Friedrich Cohnheim (1839–1884), pioneer of experimental pathology and the outstanding pupil of Rudolf Virchow, was a professor of pathology at the University of Breslau, where he published his fundamental work on the role of leukocytes in inflammation.

The fates of Cohn and Cohnheim and the city of Breslau are linked to a major event in the history of biology: in 1876, it was at Cohn's world-famous Institute of Plant Physiology that Robert Koch demonstrated the infectivity of anthrax bacilli, witnessed by Cohnheim. The pioneering studies of Koch on anthrax were published in *Contributions to the Biology of Plants*, the journal founded by Cohn in Breslau, along with many other major papers in bacteriology.

Edgar Pick

*Julius Friedrich Cohnheim—Minerva Centre of
Phagocyte Research, Sackler School of Medicine,
Tel Aviv University, Tel Aviv 69978, Israel*

Speaking of learning

How do we acquire our marvellous facility for expressing ourselves in words?

Pathways to Language: From Fetus to Adolescent

by Kyra Karmiloff and

Annette Karmiloff-Smith

Harvard University Press: 2001. 272 pp.

\$27.95, £19.50

Massimo Piattelli-Palmarini

For most people, their greatest intellectual feat is the acquisition of their native language. The virtual universality of individual success should not mask the awesome complexity of the task. Any five-year-old can tell the difference between "It rains, but it does not snow" and "It rains, but it does not matter", instantly spotting the elusive words to which the impersonal pronouns refer. There are countless examples of subtle and complex linguistic expressions that are rapidly mastered by small children, in any language or dialect.

Linguists and developmental psychologists have bravely accepted the challenge of explaining how this tremendous feat is achieved. Some have also written extended accounts for the educated but non-scientific public. Semi-popular books on children's acquisition of language already constitute a whole literary genre. In this embarrassment of riches, what can Kyra Karmiloff and Annette Karmiloff-Smith offer that is new? The answer — a not altogether appealing one — is their own peculiar view of how children master their mother tongue so well, in so short a time.

Annette and her daughter Kyra, both professional child psychologists, deliver a downbeat message formed of an eclectic embrace of many theories and levels of analysis, and a distrust of theorizing, that is entirely devoid of the excitement generated by the newer theories of language. Given that this book is intended for the general public, their message should be received with reservations as to both its content and style.

After dutifully detailing the marvel of language acquisition, the authors look at the main current theories in such areas as linguistics, psycholinguistics, developmental psychology and sociolinguistics. These topics are treated competently and concisely, if somewhat tediously. The authors then discuss the art of storytelling, in excruciating detail, describing it as "a milestone" of late language development.

Readers who have been excited, puzzled and at times outraged by the novel facts and explanations in Steven Pinker's *The Language Instinct* (William Morrow, 1994) are now presented, for the price of an occasional yawn, with sedate alternative interpretations of the same stunning data.



Mother tongue: natural ability suggests the existence of language-learning modules in the brain.

For instance, children with Williams syndrome have a barely measurable general intelligence and require constant parental care, yet they have an exquisite mastery of syntax and vocabulary. They are, however, unable to understand even the most immediate implications of their admirably constructed sentences. Karmiloff-Smith has for years strongly opposed any separation of language from general intelligence, and her book tries to dampen our amazement at this phenomenon. Errors of syntax are evident in such children if you probe more deeply, the authors say. Moreover, good auditory memory, rather than a separate and intact grammar module, may account for such observations. The authors suggest similar explanations — invoking auditory defects, for example — for other language disorders. One of these, the well-known syndrome called specific language impairment, is, as its name suggests, generally considered to be quite specific to language, though not by the present authors.

Other examples of disorders that imply the existence of a specific module for language learning are similarly turned on their heads. According to the authors, the causes are at least partially environmental, not narrowly genetic, and what is compromised in all these deficits is a lot more than just language. The disappointing message is that no clear-cut case can be made for the idea that linguistic capacities are specific and autonomous. And further, no conclusions from studying Williams syndrome or other pathological cases can be applied to the patterns of linguistic development in normal children. Moreover, the authors are keen to stress that

linguistic deficits in affected children may be partly a result of an unwitting, anomalous linguistic input from their parents.

After an excursion into the main laboratory techniques for investigating children's acquisition of their mother tongue, we are finally ready to receive the central message: no extant theory can, by itself, explain the process of language acquisition; only the combined force of all the disparate theories can do it justice. This is a bit like saying that the complexity of chemical reactions can only be explained by a combination of the quantum theory of chemical bonds and the theory of phlogiston.

Not surprisingly, this programmatic embrace of diverse theories and the pallid combination of so many factors cannot hold the reader's interest. I do not advocate the triumph of entertainment and elegance over scientific truth, but this is a book intended for a wide circulation. Few general readers will enjoy such a massive dose of disillusionment, even if it were — but it is not — scientifically well supported. An unrelenting pull towards common sense, back to granny's basic ideas of how children learn language, and the watering down into a generic, socio-pragmatic wash of many novel hypotheses that have revolutionized the whole field of linguistics needlessly spoil the fun. Pinker's linguistic crime stories are allegedly explained away. All that remains is an amalgam of mundane observations and general cognitive mechanisms that have been known for years, slightly pepped up.

Karmiloff-Smith is known to distrust those in her profession whom she considers

to be arid formalists and unbridled theory-builders. She and her daughter tell us that Noam Chomsky, Steven Pinker and hordes of other innatist, structure-oriented investigators have got it all wrong: "This type of reasoning imposes the theory onto the data, rather than letting the data constrain the theory, as good science would dictate." One might retort that few scientific revolutions would have been possible if physicists, chemists and biologists had followed this austere and stultifying precept. And we would surely not have had the so-called cognitive revolution, which was largely spurred on by daring theoretical leaps into the ultimate nature of language, the human mind and the modular organization of natural intelligence. The same can indeed be said for the discovery of innate components of the language faculty that starkly separate a child's linguistic capabilities from general intelligence, social conventions and the practical uses of language.

The authors neither totally reject nor endorse these new ideas; they just keep preaching prudence, prudence, prudence. Several passages advise researchers in the field to proceed with the utmost caution in theorizing about children and language acquisition. As a result, we have a book that appears condescending and uniquely optimized for the training of graduate students intending to work in this field — provided they agree with the authors' ultra-cautious approach. ■

Massimo Piattelli-Palmarini is in the Department of Linguistics, University of Arizona, Tucson, Arizona 85721, USA.

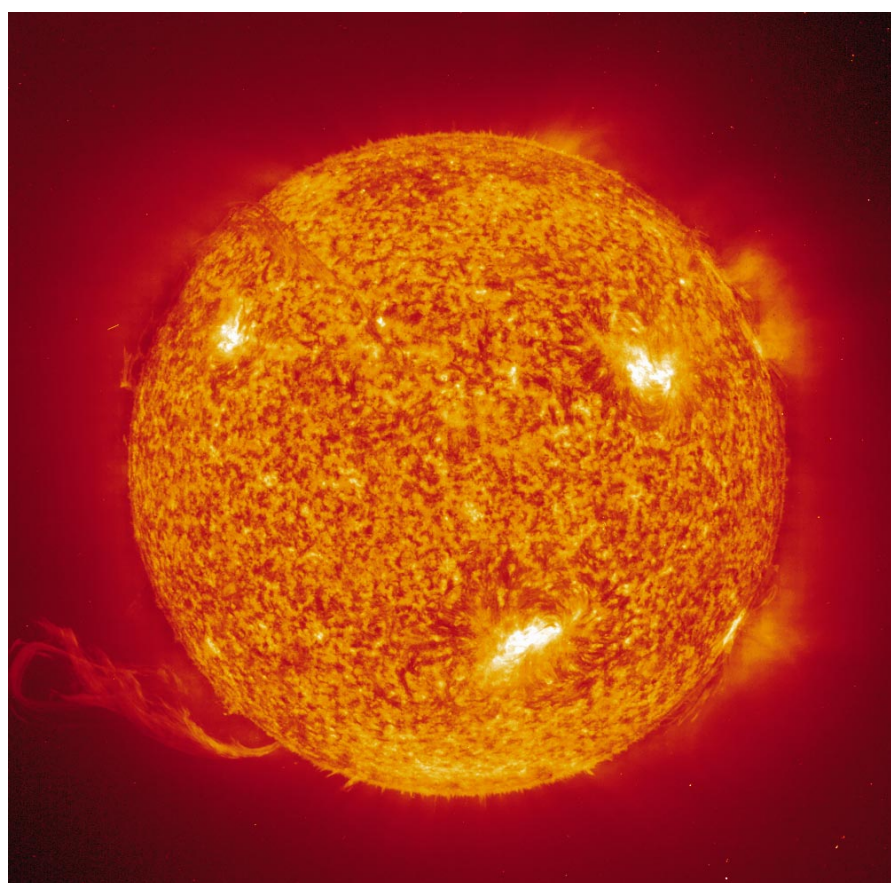
The birth of matter

The Magic Furnace: The Search for the Origins of Atoms
by Marcus Chown

Oxford University Press: 2001. 240 pp. \$25

Hans A. Bethe

The Magic Furnace tells the story of the discovery of the birth of atoms inside stars. Is it a good book? When I was about halfway through, I was doubtful. But the second half, which discusses the build-up of atomic nuclei, is very good. Marcus Chown is happiest when recounting the personal stories of scientists and their achievements. The account of Gustav Kirchhoff's discovery that each element has a characteristic spectrum, and his identification of the elements in the spectrum of sunlight, makes enjoyable reading. And the book reaches its climax with the work of Fred Hoyle and Willy Fowler, who discovered that some atoms were formed in the Big Bang



Fiery origins: atoms continue to be forged in the hot interiors of stars.

and others continue to be formed in stars.

The trouble with the first half of the book is that the author does not distinguish clearly enough between new and important findings and those that are well known. Thus, in discussing Arthur Eddington's work on stars, Chown writes the "gas is to generate a high pressure"; it has been known since Robert Boyle's time (1662) that gas has a pressure proportional to its density and temperature. Eddington would not have started his work on stars had he not known this.

Chown describes the carbon–nitrogen cycle and the proton–proton chain, in which hydrogen is converted to helium. But although he tells us that the carbon–nitrogen cycle is sensitive to temperature, we are not told that the temperature at the centre of a star (which can be as much as 40 million degrees kelvin) depends on the material that makes up its bulk. Only much later does the author tell us that stellar material is not like the material on Earth, but is predominantly hydrogen; this reduces the computed central temperature of the Sun from 40 to 13 million degrees kelvin.

The Magic Furnace ends with an informative account of how atoms are made in the extremely hot material in the Big Bang, and in the less hot interiors of stars. It explains why the Big Bang made only helium, and why there is about 25% (by weight) of helium in the Universe. We discover why

there is an abundance of atoms up to the atomic weight of iron, but fewer of those of higher atomic weight, and how the latter are formed in stars by the addition of neutrons to existing nuclei, one by one. And finally we are told how the atoms created in stars escape into the interstellar medium to form new stars. ■

Hans A. Bethe is emeritus professor of physics, Cornell University, Ithaca, New York 14853-5001, USA.

Geometry down the centuries

Euclid's Window: The Story of Geometry from Parallel Lines to Hyperspace

by Leonard Mlodinow

Free Press: 2001. 306 pp. \$26

Jeremy Gray

As its subtitle suggests, *Euclid's Window* aims to take the general reader from elementary ideas of euclidean geometry to Einstein's theory of general relativity and beyond, in this case to the theory of superstrings and their generalizations. Leonard Mlodinow takes a chronological route, describing the mathematics as he goes. So we read a bit

about euclidean geometry, cartesian coordinate geometry, non-euclidean geometry and special and general relativity in a way that minimizes the technicalities. In the final sections of the book, Mlodinow's own background in physics enables him to introduce some of those involved, such as John Schwarz, Brian Greene and Ed Witten.

Very few proofs or arguments are given. Whether the topic be mathematics or physics, all are presented as descriptions. This leaves the mathematics floating rather more freely than mathematicians might like, because mathematicians are used to explanations and arguments. But a style often found in popular accounts of modern physics may well reach a larger audience, and the interested reader can find the proofs in other books. The information given here is sufficiently accurate, although the account of the Michelson–Morley experiment (which showed that the speed of light does not depend on the direction of transmission) is rather too imprecise. And there are aspects of the book that will not satisfy everyone.

The deliberately light treatment given to the mathematics makes it almost impossible to connect up the topics in a satisfactory way. Real connections do exist, but Mlodinow has judged them too hard to describe, and the results may not hang convincingly together. The book's determinedly jokey tone becomes less than attractive after a time, and it comes unstuck when attempting to treat the historical material. There are many historical passages, sometimes taken from better recent accounts. But Mlodinow has opted for the storyteller's omnipotent eye: we are at Gauss's bedside as he dies, we are privy to people's thoughts and offered gossip (Schrödinger as the Don Juan of modern physics). Inevitably, there are errors along

with insights. But the relentlessly flippant tone undermines the seriousness of the subject. "Iced by the Snow Queen" is not the richest way to describe Descartes' death from pneumonia at the court of Queen Christina of Sweden. And there are many such touches that rob the account of the complexity required to be adequate historically.

Euclid's Window will convince any reader that important material has come out of the centuries of geometry. But readers of *Nature* may want to start with something more substantial, such as Brian Greene's *The Elegant Universe* (Vintage, 2000), which starts with Einstein and is a popular book with much more to say.

Jeremy Gray is at the Open University,
Milton Keynes MK7 6AA, UK.

Universal knowledge

Encyclopedia of Astronomy and Astrophysics, Vols 1–4
editor-in-chief Paul Murdin
Institute of Physics & Nature Publishing Group: 2001. 3,670 pp. £450, \$650 (printed set); £130, \$200 (online only); £650, \$850 (bundle)

Theodore P. Snow

Astronomers are fond of referring to their field as the "science of everything", and often boast of their own generality and, well, universal knowledge. Imagine, then, the scope of an encyclopaedia of astronomy and astrophysics — an encyclopaedia of everything. The creation of such a compendium is the task taken on by the publishers of the

Encyclopedia of Astronomy and Astrophysics (EAA). The result is, predictably, massive — but successful, at least for a specialized audience.

The *EAA* is available in print, in four large volumes, and, for subscribers, on the Internet (<http://www.ency-astro.com>; non-subscribers can access special features as well as the opening lines of any entry in the encyclopaedia — but they must become subscribers in order to access the full articles). The goals and intended uses of the *EAA* are set out in Paul Murdin's introduction. The work is aimed at "an exceptionally broad audience" — professionals and students in astronomy, professionals in related fields (such as biology or meteorology), and non-specialists seeking specific information about definitions or biographical information.

Topics are generally treated at quite a high and fairly technical level, in keeping with the aim of providing a useful resource for students and professionals. The content is a mixture of encyclopaedia, dictionary and biographical listing. Short entries defining terms or giving potted biographies are written by staff editors, and unsigned; longer articles on specialized topics are authored by leading researchers. Accuracy is ensured by the expertise of the contributors, who form an impressive who's who of modern astronomy and astrophysics.

The encyclopaedia's coverage is broad and thorough, and in some cases the depth is astonishing. In many areas, multiple entries by different contributors present different facets of the same topic. These create a complete picture that is extremely valuable, even to a specialist. There are no fewer than eight entries by different authors on 'Chromospheres', for example, each devoted to a

New in paperback

Defenders of the Truth: The Sociobiology Debate

by Ullica Segerstråle

Oxford University Press, £8.99

"Segerstråle's goal was to understand the factors that influenced the course of this scientific movement [sociobiology]. On what basis do the sociobiologists as well as their opponents evaluate sociobiology?" David Hull (*Nature* 407, 673–674; 2000)

Consciousness: How Matter Becomes Imagination — previously called A Universe of Consciousness: How Matter Becomes Imagination

by Gerald M. Edelman & Giulio Tononi

Penguin, £8.99

"Explaining consciousness has become the Holy Grail of modern neuroscience. Any reckoning on who has found the true path is

surely premature. Nevertheless, the account of consciousness provided by Edelman and Tononi is certainly highly plausible and can be recommended as one of the most ambitious accounts around." Raymond J. Dolan (*Nature* 407, 450–451; 2000)

The Voice of Genius: Conversations with Nobel Scientists and Other Luminaries

by Denis Brian

Perseus, £14.99

Stepping Stones: The Making of Our Home World

by Stephen Drury

Oxford University Press, £12.99

Charles Darwin's Beagle Diary

edited by R. D. Keynes

Cambridge University Press, £14.95, \$21.95

Darwin's Ghost: The Origin of Species

by Steve Jones

Ballantine, \$15

Lucy's Legacy: Sex and Intelligence in Human Evolution

by Alison Jolly

Harvard University Press, \$18.95, £12.95

The Sun, the Genome and the Internet

by Freeman J. Dyson

Oxford University Press, £6.99, \$10.95

Bold Science: Seven Scientists Who Are Changing Our World

by Fred Anton

W. H. Freeman, \$14.95



Star-gazing made easy

Star trails over the dome of the William Herschel Telescope on La Palma in the Canary Islands, obtained using time-exposure photography. The image is taken from

Stars & Planets (Princeton University Press, \$19.95, £12.95; pbk), a guide to the night sky, written by Ian Ridpath, with sky charts and diagrams by Wil Tirion.

subfield of the topic. One therefore gets a clear impression of the complexity of the field while also gaining a broad understanding. Some overlap and repetition are unavoidable, but this strategy ensures full coverage and representation of different viewpoints. Although anyone can quibble about the selection of topics (for example, I found no entry for one of my favourites, the diffuse interstellar bands), and it is easy to criticize the somewhat random selection of people in the biographical listings, on the whole the coverage is broad and fair, and will undoubtedly improve with time and future revisions.

The *EAA* is thoroughly cross-referenced. The short, unsigned entries include references to other related entries, while the contributed articles provide literature citations for the reader to investigate a topic more deeply. In the web edition some of the litera-

ture references appear as live links, but in most cases these are to subscription-based online resources to which the user must apply separately in order to gain access, limiting their value.

The printed version includes mainly black-and-white illustrations. There are also a few scattered four-colour inserts, but these are mostly useless and could be omitted in future editions—even a novice web user can find similar or better colour images online in a matter of moments.

The *EAA* website is well organized, utilitarian and, once a few bugs have been ironed out, quite efficient and useful. The basic content mirrors the printed encyclopaedia, but the online version has several advantages. Content will be updated (quarterly) to reflect new discoveries and to rectify errors, and access is more rapid.

Although the website may seem slow to experienced Internet explorers, it is much quicker to find an entry online than to look it up in a four-volume set of books. Licensed users can do online standard and specialized searches by subject, title, author and even illustration. And readers can set up their own internal 'bookmark' file, collecting links to articles of interest. In time, people with common interests in specific topics will be able to participate in user groups (this feature was not yet active at the time of writing).

Non-subscribers receive regularly updated 'breaking news' items, connections to newsworthy articles in *Nature*, and daily updates on the current status of solar activity and 'space weather', as well as an "Article of the week" and "Picture of the week".

Although useful for the lay reader, this compendium will undoubtedly be most used and of greatest value to the professional and educational communities. Subscription costs are significant, so users may want to organize institutional or group licences. ■

Theodore P. Snow is at the Center for Astrophysics and Space Astronomy, University of Colorado, 389 UCB, Boulder, Colorado 80309-0389, USA.

More astronomical definitions

Cambridge Dictionary of Astronomy

by Jacqueline Mitton

Cambridge University Press, £45, \$74.95 (hbk); £16.95, \$27.95 (pbk)

New Journals

This year, *Nature's* annual new journals review supplement will appear in the 18 October issue. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1999 and published at least four separate numbers by the end of June 2001.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language must be English.
- Deadline for submission is 5 July.

For each eligible title, please send at least four different issues (the first, the most recent and any two others), together with full details of subscription rates, to: Isobel Flanagan, *Nature*, The Macmillan Building, 4 Crinan Street, London N1 9XW, UK. Tel: +44 (0)20 7843 4542. Fax: +44 (0)20 7843 4596/7. e-mail: i.flanagan@nature.com

Risk and uncertainty

At the frontiers of science, we don't always know what may happen.

Robert May

The science we encounter at school deals mainly with crisp certainties, such as newtonian predictions of planetary orbits or the underlying reasons for arranging elements in the periodic table. Otherwise, science appears as rather tedious lists of names — of chemicals or plants and animals — in the trivial guise of a television quiz. This situation is as understandable as it is unfortunate: understandable because such certainties, derived from yesterday's research, are naturally easier to teach; and unfortunate because the implied equation of science with certainty is misleading.

The 'already understood' is thus the limit of most people's acquaintance with science. Insofar as uncertainty enters, it is because the underlying laws, although known, manifest themselves in a sufficiently complicated way that only probabilistic statements can be made. Examples range from roulette wheels to the chaotic dynamics of local weather prediction. Given this conception, or misconception, of what science is all about, it is not surprising that people expect scientists to give them clear and unambiguous advice when new and worrying problems appear.

Difficulties arise when the uncertainties in scientific advice to policy-makers are not caused by probabilistic predictions, but rather derive from a fundamental lack of understanding of new phenomena at or beyond the frontiers of present knowledge. Having met science as 'the known', many people balk at scientists saying "I do not know, but here is a reasonable guess". Zealots who preach a particular answer with unfounded but unshakeable belief often make these problems worse.

Medical advances have not only lengthened life expectancy, but have also reduced its variance in the developed world; we now regard three score years and ten as a basic right, and look for someone to sue if a relative does not get them. Such expectation of longer and healthier lives results, I think, in our often worrying about relatively minor

risks in ways that an earlier age would find incomprehensible. Today, some parents are just as concerned about conjectured risks from vaccinations as an earlier generation of parents was about the very real risks of childhood infections such as poliomyelitis. By the same token, many people happily accept risks where they have the illusion of control (driving cars, for example) whereas they flatly reject risks of vastly lower probability, which, however, have a 'miasmatic' quality, such as the perceived risks from radiation. Such subjective perceptions create their own realities, but they frequently ignore objective or 'scientific' approaches to risk analysis.

Whether scientific advice is a confident prediction, a statistical distribution of outcomes, imaginative guidance when we are scientifically ignorant, or nothing more than a sharper set of questions to guide us in the fog, the results will necessarily play out in an arena shaped by public perceptions. Public attitudes to risks can be hugely affected by the emotional colour of particular words. In Britain, the public-health benefits of irradiated food have not been realized as they have in many other countries, essentially as a result of public reaction to the word 'irradiated', with its misleading echoes of nuclear radiation. Appreciating this point, the medical community delivered the benefits of 'nuclear magnetic resonance' techniques, without evoking anxiety, by renaming it 'magnetic resonance imaging'.

Conversely, there are strong indications that heavy intake of vitamin B₆ can be harmful, but proposed restrictions caused an outcry from people who felt that a vitamin must, by definition, be beneficial. As another example, I share many of the environmental ideals of the organic farming movement, but I find it fascinating how the word 'organic' has been cut free from any biological moorings and made emotive, synonymous with 'good' — whereas many 'organic' plants contain naturally occurring toxins.

There are substantial questions that need answering in relation to genetically modified foods, especially in their potential for further intensification of agriculture with consequences for the countryside and the plants and animals in it. But much of the discussion has been conducted in terms of the emotional loading on words such as 'mutant' or 'genetic' — not to mention "Frankenstein foods" — in ways that create feelings divorced from factual underpinning. Today in Europe, many people feel apprehensive about mixing genetically engineered soya beans with those produced as a result of generations of artificial selection,



Uncertain outcome: even with sophisticated data collection and computation, weather prediction is still probabilistic.

NEWSMAKERS

whereas such mixing causes little concern in the United States. It is easy to forget that, when the underlying gene-splicing technology was first emerging in the 1970s, it was Americans but not Europeans who were apprehensive. At one major debate of that time in Cambridge, Massachusetts, one speaker began by intoning "nicotine, cigarettes" and then "gene splicing, genetic engineering", observing that although the former words have little emotional resonance, the latter cause shivers. But the first set of words stand for demonstrable harm on a large scale, whereas the latter do not.

The past century saw great advances in scientific understanding, applied with good intentions to make life better. But we now begin to see unintended adverse consequences: climate change and diminishing biological diversity. In the new century, society needs to do a better job of deciding what kind of world it wants to make with the opportunities science offers, rather than just letting things happen. This is a debate about values, with science having no special voice except in factual clarification of possibilities and constraints. But the task is as hugely difficult as it is hugely important. And a large part of the difficulty lies in the uncertainties that are an inseparable part of science at the frontier. It helps to recognize, and explicitly acknowledge, these uncertainties; it hinders further to cloud the uncertainty with emotionally coloured language. ■

Robert May is President of the Royal Society, Carlton House Terrace, London SW1Y 5AG, UK, and former chief scientific adviser to the UK government.

Public attitudes to risks can be hugely affected by the emotional colour of particular words.

The zombie within

Christof Koch and Francis Crick

To what extent are we conscious of everything going on in our brains? Nietzsche and Freud popularized the notion of the unconscious as a realm of the mind that controls human behaviour but is not itself accessible to conscious introspection or knowledge. By 'unconscious' we mean any neuronal activity that does not give rise to conscious sensation, thought or memory. Although many of Freud's ideas, involving penis envy, the Oedipus complex, the Id and other fanciful creations, are mere myths that lack objective standing, science has provided credible evidence for the existence of sensorimotor systems in the primate brain that function in the absence of consciousness.

Many mammalian brain systems perform complex yet routine tasks without direct conscious input. Melvyn Goodale and David Milner described the brain parts responsible for this as 'online systems', by analogy with software that processes information in real time. Such systems can deal with certain commonly encountered situations automatically, which is why we call them 'zombie' agents. One can become conscious of the actions of one's own zombie, but usually only in retrospect.

The best evidence comes from studying dissociation of 'vision for perception' and 'vision for action' in both healthy humans and

patients. For instance, an experimental subject, sitting in the dark, stares at a single light source and points a finger at it. Suddenly a new light appears in the periphery and the observer has to move his/her eyes and finger rapidly towards this target. While the eyes are in transit, the light is moved a little bit to the left or right. Neither eyes nor finger has a problem in correcting for this and both end up right on target. Yet the subject does not see the extra target motion, even though the oculomotor system automatically corrects for it. All that is perceived is a flashing target and the eyes and hands making a movement towards it.

Other instances of online systems are those that adjust for body posture on the basis of visual or vestibular cues, or that control the way we shape our hand and fingers when we reach out to pick something up. Rigorous experiments involving forced choices between two alternatives show that we can respond to scary pictures of snakes or spiders even when not consciously aware of them.

Within the clinical domain, some neurological patients show very selective deficits. A key element of the Milner–Goodale argument is patient D.F., who cannot visually recognize objects or shapes since suffering carbon-monoxide-induced anoxia. Ask her whether an elongated slot is vertical or horizontal and she has to guess. Yet when she 'posts' a letter into this slot, she effortlessly rotates her hand into the appropriate position. However, when she is shown the slot, the light is then turned off and she has to wait a few seconds before executing the hand movement, she fails. This might imply that her online system does not have access to explicit memory. When she reaches out to pick up a pencil or a glass, she scales her handgrip accordingly, as do normal subjects. Yet she denies seeing the object.

Even more spectacular cases of zombie behaviour can occur in patients with complex partial seizures and in sleepwalkers. Both involve complex yet relatively stereotypical motor patterns: wandering around, moving furniture and even driving cars. This automatic behaviour follows an internal programme that can be influenced by the environment (for example, when sidestepping an obstacle). In general, neither the epileptic patient nor the sleepwalker responds to commands or remembers anything later. The simplest interpretation is that, although consciousness is shut down by the partial seizure or by deep sleep, enough of the forebrain remains active to subserve online systems. Both syndromes accentuate the difficulty of assessing the degree of consciousness in the absence of either explicit recall or language.

Online systems are fast, outpacing consciousness. Anecdotal evidence and

Unconsciousness

"There is credible evidence for the existence of sensorimotor systems in the primate brain that function in the absence of consciousness. So why bother with consciousness?"

psychophysical research emphasize rapid and effortless behaviour that predates consciousness. This is particularly true of the highly practised and ritualized sensorimotor activities that humans love, such as rock-climbing, fencing and dancing. Mastery of these requires a surrendering of the conscious mind, allowing the body to take over.

The hallmarks of a zombie system are stereotypical, limited sensorimotor behaviour, and immediate, rapid action. Its existence raises two questions. First, why aren't we just big bundles of unconscious zombie agents? Why bother with consciousness, which takes hundreds of milliseconds to set in? It may be because consciousness allows the system to plan future actions, opening up a potentially infinite behavioural repertoire and making explicit memory possible.

Second, what is the difference between the neuronal pathways that subserve zombie agents and the neural networks that give rise to specific, conscious perception? Both probably involve the cerebral cortex and the thalamus. Are they based on activity in different subsets of neurons, segregated according to brain areas? Could the neuronal correlates of consciousness correspond to a cell type, intermixed with other types that are responsible for unconscious behaviour? Or might the difference be the type of neural activity involved? We hypothesized earlier that consciousness involves synchronized firing of neurons at the millisecond level, whereas uncorrelated firing can influence behaviour without generating that special buzz in the head.

Could mutation of a single gene turn a conscious animal into a zombie? If so, what test would show that they are unconscious? Tracking down the neuronal correlates of consciousness in humans, monkeys and mice should illuminate the central mystery of how neural activity in specific feedback circuits gives rise to subjective states. ■

Christof Koch is in the Divisions of Biology and of Engineering and Applied Science, Caltech, Pasadena, California 91125, USA. Francis Crick is at the Salk Institute, La Jolla, California 92037, USA.

FURTHER READING

<http://www.klab.caltech.edu/cns120>

Crick, F. *The Astonishing Hypothesis* (Charles Scribner's Sons, New York, 1994).

Milner, A. D. & Goodale, M. A. *The Visual Brain in Action* (Oxford Univ. Press, Oxford, 1995).



Offline: Lady Macbeth goes walking in her sleep.



Non-stick water

L. Mahadevan

Adding a simple powder to a drop of water gives it remarkable properties: the powder-coated drop no longer sticks to surfaces, and moves by rolling, much as a solid sphere would.

Much to the consternation of adults, a broken mercury thermometer is a source of delight to a curious child absorbed by the spectacle of the silvery beads that elude capture. The epithet 'quicksilver' aptly describes the droplets, which seem to roll rapidly on the surface much like a solid marble. This unusual behaviour elicits a host of questions that are relevant to scientists in diverse fields: from chemists and engineers interested in the dynamics of drops, to astronomers interested in the stability of rotating stars and planets. How and when can liquid droplets actually roll on a surface? How fast can they move? Can they be controlled? And to what end?

On page 924 of this issue¹, Aussillous and Quéré describe the results of experiments that begin to address these issues. One motivation for these studies was provided by the need to understand the wetting properties of surfaces with different affinities for liquids. These include hydrophobic (water-repellent) surfaces, which are important in industrial processes, such as metal refining, and biological processes, such as cell motility. Traditional methods of creating these surfaces rely on chemical treatments; a familiar example is wax paper on which water will form beads.

But none of these techniques is particularly robust or long lasting, except in very clean environments. Moreover, the contact angle between the water-repellent surface and the water droplet, which characterizes the degree of hydrophobicity, cannot be increased much beyond 120° using purely chemical means. To make a super-water-repellent surface (one with a contact angle greater than 120°) requires new tricks.

In the past five years, chemical methods have been supplemented by a structural approach. On rough surfaces, contact angles of up to 175° have been achieved with water

droplets². On these super-hydrophobic surfaces, the water cannot enter the microscopic chasms in the roughness because the pressure required is too high. Therefore, the droplet is in contact with only a small fraction of the solid, much like an Indian fakir on a bed of nails, and the effective contact angle becomes large.

Aussillous and Quéré¹ have ingeniously inverted these ideas. They add a drop of water to a powder of silane-coated spores of the club moss *Lycopodium*, and stir the mixture until it yields nearly spherical, composite droplets. The powder-coated droplets are non-wetting—the resulting liquid marble will even float on a pool of water. The droplet is also a soft solid: it has an elasticity that arises from surface tension, and can bounce and roll without leaking.

The properties of these liquid marbles are brought to the fore by the lack of interaction with the underlying surface. In contrast, most liquids, such as water on smooth glass, are partially wetting, and solid–liquid interactions dominate, leading to a finite contact angle (Fig. 1a). Compared with partially wet-

ting liquids, the geometry of contact is qualitatively different for non-wetting liquids because the angle of contact is 180°, and the droplet shape resembles that of a flattened sphere (Fig. 1b). Previous theoretical investigations of non-wetting droplets³ predicted behaviour that seemed unusual and controversial. But the difficulty of achieving perfectly non-wetting surfaces prevented the theory from being tested quantitatively. Now, using their non-wetting liquid marbles, Aussillous and Quéré show that these pristine soft solids do behave unusually.

Theory tells us that, for static liquid marbles, the size of the contact area is determined by the balance between gravity (which favours contact) and capillary forces (which oppose it). The geometry of contact also dictates what happens when the marbles begin to move. A non-wetting droplet has the ability to roll without slipping on a surface that is only slightly inclined, much as a solid sphere would. By contrast, partially wetting droplets can remain stuck even on a vertical window pane. When these partially wetting droplets move they slide rather than

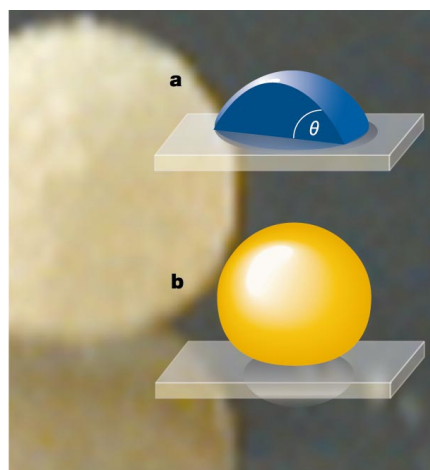


Figure 1 Wetting and non-wetting drops. a, A small partially wetting droplet on a solid surface has a finite contact angle, θ , determined by the behaviour of the vapour–liquid–solid system. The wedge-like shape of the contact zone controls both the statics and the dynamics of these partially wetting droplets⁴. b, Non-wetting droplets, such as the 'liquid marbles' created by Aussillous and Quéré¹, have an effective contact angle of 180° and resemble slightly flattened spheres. These liquid marbles, one of which is shown in the background and overleaf, have a different contact geometry from partially wetting droplets, and so have very different static and dynamic behaviours.

roll because their motion is controlled by the wedge-like geometry of the solid–liquid interface⁴ in Fig. 1a.

What happens to non-wetting droplets when they move on a slightly inclined plane? The velocity of the rolling marbles is ultimately determined by the competition between gravity (acting on the mass) which speeds them up, and internal fluid dissipation (in a region determined by the contact area), which slows them down. Theory predicts³, and experiments with liquid marbles now show¹, that larger droplets roll more slowly than smaller ones. This seemingly paradoxical result has a simple origin: larger drops have a larger contact area, which increases more rapidly than the volume (mass) as the size of the drop increases. So smaller drops should roll faster than larger ones, as long as the drops remain nearly spherical.

The behaviour of the liquid marbles is even more dramatic when they move rapidly. If the surface on which a marble rests is tilted to large angles, the droplet starts to spin rapidly and is transformed into a wheel-like disc with a thin web-like hub, reminiscent of planetary rings. At these high velocities the inertial forces due to rotation can greatly deform the drops. The possibility of such large deformations led Henri Poincaré and Lord Rayleigh to consider freely spinning toroidal shapes⁵ more than a century ago.

Later considerations suggested that such shapes are unstable⁶. In the new study¹, the presence of the inclined plane seems to be sufficient to stabilize the spinning wheel, and as soon as the drop either lifts off the plane or leaves the edge, the wheel-like shape is no longer stable and it rapidly morphs into a stable peanut-like shape. By looking closely at successive snapshots we can follow the evolution of the instability (see the image at the top of the previous page). But this still leaves open the details of the stabilization mechanism. Aussillous and Quéré's simple method for producing freely spinning drops will hopefully invigorate Earth-bound experiments on a centuries-old problem, complementing gravity-free experiments in space^{7,8}.

The ease with which these liquid marbles can be manufactured and moved, coupled with the relatively large velocities that can be achieved, makes them suitable for use in rapid, wear-free micromachines, such as electromechanical actuators and valves. Another possible application could be in lubrication, either as individual liquid ball-bearings or as a confined emulsion. But before such applications can be tackled, questions of how robust the drops are, and

the way their behaviour changes with age, need to be addressed.

Finally, we may not have to look far for situations that favour the production of these liquid marbles. The leaves of plants such as the lotus⁹ have many closely spaced, stiff, waxy protruberances, and many biological surfaces are hairy; a combination of chemistry and structure keeps the surfaces dry (and clean) by minimizing the effective area exposed to contaminants. Wet, dusty environments should also favour the formation of non-wetting droplets, which may play a role in transporting powders. Analogues of liquid marbles might also be relevant in cell biology, where adherent cells sometimes move by rolling¹⁰. As for mercury itself, alas it deceives the eye



Structural biology

Chlorophylls galore

Werner Kühlbrandt

A high-resolution crystal structure of photosystem I, part of the machinery that performs photosynthesis, reveals how an extensive array of chlorophylls uses solar energy to transport electrons.

As more and more genomes are decoded, the need to understand in detail the molecular mechanisms of the encoded proteins becomes pressing. This is particularly true for proteins found in cellular membranes, as the genes encoding membrane proteins account for roughly a quarter of most genomes. A prerequisite for understanding how a protein works is knowledge of its precise three-dimensional crystal structure, but so far such structures are available for only a few membrane proteins. However, the energy-converting proteins found in bacterial and plant photosynthetic membranes, which carry out one of the most fundamental processes in biology, have yielded to crystallography more readily than others.

Now, on page 909 of this issue¹, Jordan and colleagues report the crystal structure of photosystem I, a large photosynthetic assembly of membrane proteins and other factors, at 2.5 Å resolution. The structure is the culmination of decades of painstaking work in the authors' labs to perfect both the biochemical preparation of the complex and the quality of its crystals.

There are two different but related forms of photosystem, known as type I and type II. The most ancient and primitive photosynthetic bacteria have either one or the other, but plants and cyanobacteria have both, and use them in sequence to convert solar energy into chemical energy (Fig. 1). Essentially, photosystems use the excitation energy of

and is not perfectly non-wetting, although mixing it with a powder should probably do the trick. Nonetheless, it is evident that there is still much to marvel at in a tiny droplet, if we choose to observe it with the delight of a child's eye.

L. Mahadevan is in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK.

e-mail: l.mahadevan@damtp.cam.ac.uk

1. Aussillous, P. & Quéré, D. *Nature* **411**, 924–927 (2001).
2. Onda, T., Shibuchi, S., Satoh, N. & Tsujii, T. *Langmuir* **12**, 2125–2127 (1996).
3. Mahadevan, L. & Pomeau, Y. *Phys. Fluids* **11**, 2449–2454 (1999).
4. de Gennes, P. G. *Rev. Mod. Phys.* **57**, 827–863 (1985).
5. Rayleigh, J. W. S. *Phil. Mag.* **28**, 161–170 (1914).
6. Brown, R. A. & Scriven, L. E. *Proc. R. Soc. Lond. A* **371**, 361–377 (1980).
7. Wang, T. G. et al. *J. Fluid. Mech.* **276**, 389–403 (1994).
8. Lee, C. P. et al. *J. Fluid. Mech.* **354**, 43–67 (1998).
9. Neinhuis, C. & Barthlott, W. *Ann. Bot.* **79**, 667–677 (1997).
10. Anderson, K. I., Wang, Y.-L. & Small, J. V. *J. Cell. Biol.* **134**, 1209–1218 (1996).

sunlight, absorbed by chlorophyll, to extract electrons from a substrate and push them through the membrane in which the photosystems sit, creating an electrochemical gradient across the membrane. For photosystem I the usual substrate is a small, soluble redox protein called plastocyanin; for photosystem II, it is water. The electrochemical gradient powers the production of the energy-storing molecule ATP and, directly or indirectly, leads to the biosynthesis of nearly all organic matter on our planet.

One of the waste products of photosynthesis is oxygen, and photosynthesis by cyanobacteria is thought to have been the main reason that oxygen levels in the Earth's atmosphere rose to the point at which animal life could develop about 2 billion years ago. Since then, the photosystems embedded in the cell membranes of cyanobacteria have barely changed. These microorganisms were the direct precursors of chloroplasts, the photosynthetic organelles in land plants and algae, so the photosystems in cyanobacteria and chloroplasts are quite similar. Some species of cyanobacteria that live in hot springs have more rugged photosynthetic complexes than do chloroplasts. Jordan et al.¹ found these hardy photosystems to be particularly suitable for crystallization, and this was part of the key to their success.

The structure described by Jordan et al.¹ reveals photosystem I to be a bewilderingly complex assembly of no fewer than 11 differ-

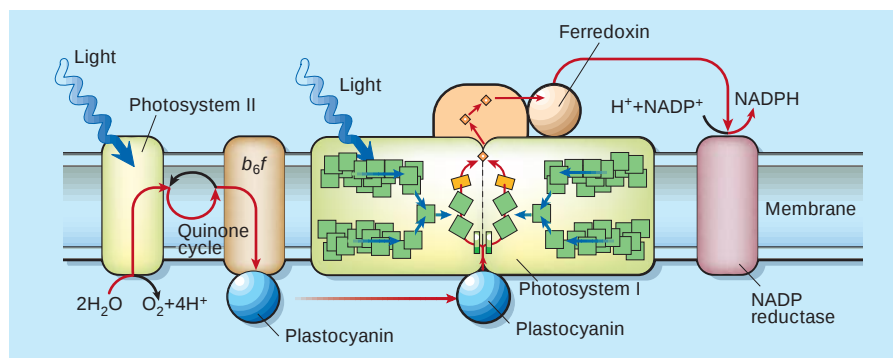


Figure 1 Flow of excitation energy (blue arrows) and electrons (red arrows) through photosystem I, as seen in the 2.5 Å structure obtained by Jordan *et al.*¹. From left, electrons extracted from water by photosystem II in the photosynthetic membrane pass through the quinone cycle and the cytochrome *b₆/f* complex (brown) to plastocyanin (blue), a soluble redox protein, which then docks at photosystem I. Solar energy absorbed by photosystem I's antenna chlorophylls (green squares) is transmitted to the central electron-transfer chain of photosystem I. The electron-transfer chain consists of six chlorophylls and two bound quinones (yellow), related by an imperfect twofold symmetry axis¹ (dotted line). Electrons from plastocyanin are transferred through either branch of the electron-transfer chain to a set of three iron-sulphur clusters (orange diamonds). The electrons are accepted by ferredoxin, a soluble redox protein (brown), and are finally used by NADP reductase to produce NADPH, which enables the cell to synthesize its molecular components. The resulting proton gradient across the membrane powers the generation of ATP.

ent polypeptides, 8 of them membrane-spanning. These proteins bind 96 chlorophylls and 31 other small molecules, all held at the precise distances and relative orientations that are required for efficient absorption and conversion of solar energy. The main components of photosystem I display near-twofold symmetry around a special pair of chlorophylls, which are at the start of an electron-transfer chain and extract electrons from plastocyanin. The same symmetry is common to all photosynthetic reaction centres, suggesting that it is essential to endow the central pair of chlorophylls with the spectral and redox properties necessary to remove electrons from a donor.

The structure¹ shows that, from the central pair of chlorophylls, electrons are passed along a chain of cofactors — through a set of four accessory chlorophylls and two redox-active phylloquinones (vitamin K molecules) to three strategically positioned iron-sulphur clusters (Fig. 1). There the electrons are accepted by ferredoxin. This small, soluble redox protein shuttles reducing power to an enzyme called NADP reductase, which fixes the electrons in a form usable for cellular biosynthesis reactions.

As a consequence of the near-twofold symmetry, this electron-transfer chain has two slightly different branches. In type II photosystems, one branch is turned off. Intriguingly, however, both branches seem to be active in photosystem I. Why might this be? The type II electron-transfer chain ends at a quinone, which acts as a lipid-soluble electron carrier, and needs to be fully reduced to a stable quinol by two electrons before it can be released into the membrane. This would take too long if both branches

were active². By contrast, the quinones of photosystem I are integrated into the transfer chain¹, handing over their electrons rapidly to the same iron-sulphur cluster, so presumably there was no evolutionary pressure to turn one chain off.

Only 6 of the 96 bound chlorophylls in photosystem I are engaged in redox chemistry and electron transfer¹. The remaining 90 act as an antenna, harvesting solar energy and feeding it into the electron-transfer chain. All except two of the antenna chlorophylls keep a respectful distance from the electron path, suggesting that excitation energy might be fed into the transfer chain through these two 'bridging' chlorophylls.

Most of the antenna chlorophylls are positioned in two layers, an arrangement similar to that first observed in the plant light-harvesting complex II³. But the frequent use of water molecules in coordinating chlorophylls in photosystem I is surprising, as is the attachment of two chlorophylls in the electron-transfer chain to sulphur-containing side chains of nearby amino acids. Small clusters of two or three chlorophylls at the periphery of the complex apparently account for the puzzling absorption by photosystem I of wavelengths at the far-red end of the solar spectrum. The coordination of one chlorophyll by the head group of a tightly bound lipid is unprecedented, and suggests that lipids have a more active role in the structure and function of some membrane proteins than we might have expected.

Earlier, low-resolution structures^{4,5} of photosystems I and II showed that the two major protein subunits of photosystem I are strikingly similar to the corresponding subunits of photosystem II — a discovery



100 YEARS AGO

The structure of the hairs of the Patagonian ground-sloth and of the living South American edentates forms the subject of an essay by Dr. W. G. Ridewood, which appears in the May issue of the *Quarterly Journal of Microscopical Science*. The most generally interesting of the author's observations are those relating to the hairs of the two living types of sloth, and the structure which permits of the growth of an alga in each. In the three-toed sloth the hair is invested with a thick extra-cortical layer. "The layer has a tendency to crack in a transverse direction, and in the cracks there come to lodge unicellular algae, to which Kühn has given the name *Pleurococcus bradypii*. The moisture of the climate in which *Bradypus* lives enables the alga to live and propagate in this curious position, and the sloth acquires a general green tint which must render it very difficult to distinguish as it hangs among the green foliage."

From *Nature* 20 June 1901.

50 YEARS AGO

One of the widely accepted tenets of plant physiology is that photosynthesis is intimately associated with the integrity of green cells. Disruption or even injury to the intact cell leads to the cessation of the process. This view has been based on the failure, for almost three-quarters of a century, of the many attempts to reproduce photosynthesis outside the living cell. The early experiments have often sought to attain an objective which has been aptly characterized as alchemistic in the light of modern knowledge: the performance by broken or dead cells or even by chlorophyll solutions of the process of photosynthesis *in toto*, that is, the utilization of light energy for the reduction of carbon dioxide and the simultaneous evolution of oxygen. However, these early attempts were not wholly negative. Although complete photosynthesis was never reproduced *in vitro*, some of the early experiments of Haberlandt, Ewart and Molisch indicated that certain cell preparations retain a limited capacity for oxygen evolution under the influence of light. The full significance of these observations did not become apparent until Hill showed by independent biochemical methods that chloroplasts removed from living cells retain for an appreciable period of time that part of the photosynthetic apparatus which is responsible for the evolution of oxygen.

From *Nature* 23 June 1951.

that led to the stunning realization that all photosystems are basically alike and must have the same evolutionary origin. A more detailed comparison of the two photosystems should soon be possible: we now have Jordan *et al.*'s high-resolution structure¹ of photosystem I, and the structure of cyanobacterial photosystem II, currently at 3.8 Å resolution⁶, is being improved and refined. Together, the two structures will provide complete insight into the workings of these fascinating solar-energy converters. ■

Werner Kühlbrandt is in the Department of Structural Biology, Max Planck Institute of Biophysics, Heinrich-Hoffmann-Strasse 7, 60528 Frankfurt am Main, Germany.

e-mail: kuehlbrandt@mpibp-frankfurt.mpg.de

1. Jordan, P. *et al.* *Nature* **411**, 909–917 (2001).
2. Deisenhofer, J. & Michel, H. *Science* **245**, 1463–1473 (1989).
3. Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. *Nature* **367**, 614–621 (1994).
4. Schubert, W. D. *et al.* *J. Mol. Biol.* **272**, 741–769 (1997).
5. Rhee, K. H., Morris, E. P., Barber, J. & Kühlbrandt, W. *Nature* **396**, 283–286 (1998).
6. Zouni, A. *et al.* *Nature* **409**, 739–743 (2001).

Astronomy

Giants in the asteroid belt

Derek C. Richardson

Models of Solar System formation have always had a hard time explaining the mysterious asteroid belt. By injecting new life into an old myth about the origin of asteroids, astronomers may now have the answer.

The asteroid belt consists of small bodies that orbit the Sun between Mars and Jupiter. On the bicentennial of the discovery of the first and largest asteroid, Ceres, by Italian astronomer Giuseppe Piazzi, it is surprising how little we understand about the history of the asteroid belt. A common misconception is that asteroids are the remains of a large planet that mysteriously exploded long ago. Today there is hardly enough material in the asteroid belt to make a small moon, let alone a planet. So something must have happened either to prevent the formation of a planet between Mars and Jupiter, or to clear out any material that accumulated there.

Typically the blame is cast on Jupiter, our largest and therefore most gravitationally influential neighbour. But exactly how it does this is a matter of some debate, because at the moment Jupiter affects only a few narrow regions of the asteroid belt. In an article in *Meteoritics and Planetary Science*, John Chambers and George Wetherill¹ paint the clearest picture yet of what may have happened. Surprisingly, they find that there may once have been Earth-sized planets in the asteroid belt, but that a combination of complex dynamics and chance events conspired to remove them.

The 'planetesimal hypothesis' of planet formation², a modern-day version of theories proposed by Kant and Laplace in the eighteenth century, states that the planets formed when smaller bodies collided and stuck together. These small bodies, called planetesimals, were part of the Solar nebula — a disk of gas and dust swirling around the Sun. In the inner regions of the Solar nebula, only non-volatile material survived to form the rocky Earth-like planets. But in the outer regions, where temperatures were cool enough to allow hydrogen-bearing

compounds to solidify, the giant gas planets formed. The present-day masses of all the planets can be explained if the mass per unit area of the Solar nebula decreased smoothly with distance from the Sun — apart from the asteroid belt^{3,4}, that is, where there is a 1,000-fold drop in the amount of matter. It seems unlikely that this deficit was present at the outset in an otherwise smoothly distributed nebula, so most theories propose that the asteroid belt was cleared later on.

In the Chambers and Wetherill model¹, planets began to form throughout the disk, including the primitive asteroid belt, where a planet-sized body could form within one million years⁵. But the large amount of volatile material around Jupiter meant that Jupiter grew rapidly, eventually perturbing the asteroidal planets. These perturbations are caused by 'resonances' between the orbits of the giant planets and young asteroidal planets. A resonance arises when the orbital periods of two bodies are whole-number ratios of each other.

For example, an object in the asteroid belt, located 3.3 times as far as the Earth from the Sun, orbits the Sun exactly twice for each orbit of Jupiter — it is in a 2:1 resonance with Jupiter. Such an object experiences periodic momentum 'kicks' that gradually elongate its orbit, increasing the risk of it crashing into another body, including a terrestrial planet. It could also end up getting too close to Jupiter or the Sun, and so be ejected permanently from the Solar System because of a 'slingshot' effect (see ref. 6 for a review).

As Jupiter grew in mass, the orbital resonances it created became stronger. Any asteroid unlucky enough to wander into one of these resonances today would be ejected from the Solar System (or crash into the Sun) within one million years⁷. But these zones of instability are much narrower than

the existing asteroid belt. Somehow, objects in the primitive asteroid belt were pushed into these zones. Chambers and Wetherill propose that encounters between the embryonic planets in the belt were responsible, at least to begin with. As the belt became depleted, this process slowed until only two or three planet-sized objects remained. It was then a matter of chance whether the stragglers survived or scattered each other into oblivion.

To check their hypothesis, Chambers and Wetherill¹ performed a series of numerical simulations in which they varied the numbers and masses of embryonic planets, and included the effects of an already formed Jupiter. They found that interactions between the embryos and Jupiter caused more than 90% of the mass in the asteroid belt to be cleared in less than a few hundred million years. Moreover, they found that if they started Jupiter (and Saturn in a few cases) in more elongated orbits, the clearing rate increased dramatically, becoming as short as a few million years. This is because the resonance zones become wider and stronger as the orbits of the perturbing planets become elongated.

At the same time, the embryos perturb the orbits of the giant planets by siphoning off energy they need to escape from the asteroid belt, thereby causing the giant orbits to become more circular. This process is analogous to that proposed for clearing objects from the outer Solar System^{8,9}. Together these effects imply that the giant planets may have had elongated orbits when they formed. Giant planets with elongated orbits have been observed in other solar systems¹⁰, so perhaps there was insufficient material left over in the nebulae of these systems to make the giant orbits circular.

The Chambers and Wetherill theory also suggests that instability induced by resonances could cause asteroidal embryos rich in volatiles, such as water, to crash into the Earth — a process that may explain the origin of Earth's oceans¹¹. Also, in roughly one-third of their simulations, a planet did survive in the outer asteroid belt. This means it was largely a matter of chance whether our asteroid belt would harbour a planet. Although such a planet could have frustrated the beginnings of advanced life on Earth by increasing the rate of impact with planetesimals, it would probably have only delayed our progress for a short while.

Like most theories, there are certain assumptions in the model that may lead to its downfall. Most serious is the need for Jupiter to form slowly enough for planet-sized bodies to form in the asteroid belt before being perturbed¹². But the elegant way in which the theory extends the standard model for terrestrial planet formation makes it attractive. Until more simulations are performed, or until there are enough data from

other planetary systems to favour one model in particular, we must keep all alternatives in mind as we pursue our understanding of planet formation.

Derek C. Richardson is in the Department of Astronomy, University of Maryland at College Park, College Park, Maryland 20742, USA.

e-mail: dcr@astro.umd.edu

1. Chambers, J. E. & Wetherill, G. W. *Meteorit. Planet. Sci.* **36**, 381–399 (2001).
2. Lissauer, J. J. *Annu. Rev. Astron. Astrophys.* **31**, 129–174 (1993).

3. Weidenschilling, S. J. *Astrophys. Space Sci.* **51**, 153–158 (1977).
4. Lissauer, J. J. *Icarus* **69**, 249–265 (1987).
5. Wetherill, G. W. & Stewart, G. R. *Icarus* **106**, 190–209 (1993).
6. Murray, N. & Holman, M. *Nature* **410**, 773–779 (2001).
7. Gladman, B. J. et al. *Science* **277**, 197–201 (1997).
8. Fernandez, J. A. & Ip, W.-H. *Icarus* **58**, 109–120 (1984).
9. Hahn, J. & Malhotra, R. *Astron. J.* **117**, 3041–3053 (1999).
10. Marcy, G. W., Cochran, W. D. & Mayor, M. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 1285–1311 (Univ. Arizona Press, Tucson, 2000).
11. Morbidelli, A. et al. *Meteorit. Planet. Sci.* **35**, 1309–1320 (2000).
12. Kortenkamp, S. J. & Wetherill, G. W. *Icarus* **143**, 60–73 (2000).

Evolutionary biology

Seeing red in speciation

Michael J. Ryan

Mating patterns in sticklebacks have been investigated for over fifty years. The latest studies show how a complex interplay between males, females and the environment can contribute to the formation of new species.

Speciation and sexual selection are central processes in evolution. Speciation occurs when divergence between organisms becomes such that they cannot produce viable offspring, and so constitute different species. Sexual selection results from female preferences for certain male attributes in choosing a mate, and promotes the evolution of extreme male appearance and behaviour during courtship.

How might these processes interact? On page 944 of this issue¹, Janette Boughman describes how she has addressed the question by studying sticklebacks in Canadian lakes. She shows how sexual selection can drive speciation through the complex interplay between ambient light levels in different parts of the lakes, male coloration, and

female sensitivity to light in different parts of the spectrum.

Mate preference is important in both speciation and sexual selection², but it gained wide recognition only with the appearance of papers by West-Eberhard³ and Lande⁴ some 20 years ago. These papers suggested that sexual selection can cause divergence of mating preferences among populations so that individuals from nearby populations perceive one another as 'different' rather than 'the same'. If so, sexual selection could generate the reproductive isolation that contributes to bringing about new species. When viewed by one another as different, individuals don't reproduce; this is reproductive isolation.

Figure 1 Courting couple — a male stickleback in red mating livery makes a pass at a female. Inset: a male with black mating coloration.



Why would populations come to differ in their mate preferences? One possibility is 'sensory drive'⁵, a change in the female perception system that could initially be unrelated to mate choice but could have an effect on it. Although sexual selection generally promotes the evolution of conspicuous male traits, 'conspicuous' is defined by the female's perceptual system and the context in which she perceives it. We know that differences among habitats can influence signal efficacy, one example being the way in which ambient light influences the conspicuousness of visual signals⁶.

In the case of the sticklebacks studied by Boughman¹, the signal sent from male to female is throat colour (usually red), which becomes more intense during the mating season. This feature has made sticklebacks the subject of numerous studies in behavioural ecology⁷ and evolution⁸, ever since Tinbergen⁹ observed that their response to red is so strong that a stickleback in an aquarium will swim at red postal trucks driving past on the road outside.

For her study, Boughman predicted the following: first, that male signals are transmitted in a habitat-specific way; in this case the red throat is more conspicuous in some parts of a lake than others because of the ambient light. Second, she predicted that the female's perceptual sensitivity adapts to local conditions, and third, that male signals match female sensitivity. If these three conditions are met, the idea is that mate preferences are then likely to diverge in different populations and that divergence will contribute to the reproductive isolation of those populations.

To test her predictions, Boughman studied six populations of sticklebacks from four lakes in British Columbia. In these populations, male mating coloration varied from red to black (Fig. 1). She found that males appeared redder, at least to her, in habitats with less red light. (This is one of the drawbacks of the study. More quantitative measures of the area and reflectance of red coloration would have allowed direct quantification of its conspicuousness to females; see, for instance, ref. 10. Subjective rankings of colour are based on reflectance as filtered by a human visual system, presumably under very different light conditions to those experienced by sticklebacks.)

To measure the sensitivity of female sticklebacks to wavelength, Boughman used experiments in which the intensity of monochromatic light is varied to determine the threshold at which a fish ceases to follow rotating black and white stripes. She found that females in areas with less red light, such as where the water is 'tea-stained' in colour rather than clear, are more sensitive to red light, and that male signals are matched to female sensitivity. So there is

the predicted three-way correlation between sensitivity, light and colour.

Cause and effect, though, are not clear. As Boughman says, male coloration could have evolved to match female spectral sensitivity, as suggested by one hypothesis (the 'sensory exploitation' hypothesis¹¹). Or both colour and sensitivity could have evolved independently to match ambient light levels. A final possibility, which Boughman does not mention, is that the contrasting colours could initially have evolved in the males, with females then developing their spectral sensitivity to match the signal. Only an analysis of evolutionary relationships could distinguish between these possibilities.

Does divergence in male traits and female preferences contribute to reproductive isolation? In this case it does. Boughman found that females that are more sensitive to red also prefer redder males within their population, and that the probability of spawning between populations is related to the degree of divergence in male redness and female preference for red. Unlike in some other studies of sticklebacks¹⁰, however, we do not know whether the mate preferences between populations are due to overall brightness, to true colour alone, or to the contrast of colour with the background or with other body parts.

Much of the study of speciation can be distilled into trying to understand how reproductive isolation happens¹². As Boughman has shown, her approach can provide detailed knowledge of how reproductive isolation comes about in sticklebacks. Future studies might tackle the intriguing subject of the neural and genetic mechanisms underlying the changes in spectral sensitivity that mediate this female preference. Is preference determined by the absorption spectrum of the photopigment that is sensitive to long wavelengths¹³? If so, one might be able to identify changes in the structure of the pigment molecules that account for such differences, and thus be able to identify one of the genes that contribute to speciation in sticklebacks (as has been claimed for the fruitfly *Drosophila*¹⁴).

Alternatively, the differences in female preference could be related to the number of long-wavelength-sensitive cones in the retina, or to the proportion of spectrally opponent neurons in the visual system. Furthermore, mathematical models of how visual scenes are analysed, some of which have been applied to sticklebacks¹⁰, could offer further insights into why not all females see red in the same way. All in all, sticklebacks may be outstanding subjects for investigating the biology of speciation in general.

Michael J. Ryan is in the Section of Integrative Biology, University of Texas, Austin, Texas 78712-1064, USA.

e-mail: mryan@mail.utexas.edu

1. Boughman, J. W. *Nature* **411**, 944–948 (2001).

2. Cronin, H. *The Ant and the Peacock* (Cambridge Univ. Press, 1991).
3. West-Eberhard, M. J. *Proc. Am. Phil. Soc.* **123**, 222–234 (1979).
4. Lande, R. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725 (1981).
5. Endler, J. A. *Am. Nat.* **131**, S125–S153 (1992).
6. Marchetti, K. *Nature* **362**, 149–152 (1993).
7. McPhail, J. D. *Can. J. Zool.* **71**, 515–523 (1995).
8. Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. *Science* **287**, 306–308 (2000).
9. Tinbergen, N. in *Psychobiology, the Biological Bases of Behavior*:

Readings from Scientific American 5–9 (Freeman, San Francisco, 1952).

10. Baube, C. L., Rowland, W. J. & Fowler, J. B. *Behaviour* **132**, 979–996 (1995).
11. Ryan, M. J. *Oxford Surv. Evol. Biol.* **7**, 157–195 (1990).
12. Coyne, J. A. & Orr, H. A. *Phil. Trans. R. Soc. Lond. B* **353**, 287–305 (1998).
13. Yokoyama, S. & Yokoyama, R. *Annu. Rev. Ecol. Syst.* **27**, 543–567 (1996).
14. Ting, C.-T., Tsaur, S.-C. & Wu, C.-I. *Proc. Natl Acad. Sci. USA* **97**, 5313–5316 (2000).

Condensed-matter physics

Nickel probes superconductivity

Michael E. Flatté

Magnetism usually destroys superconductivity, but a magnetic nickel atom inserted into a high-temperature superconductor has surprisingly little effect on its local environment.

The unusual properties of superconductors arise from the coherent behaviour of electrons when they flow together in pairs. Two key properties for electrons in a superconductor are their negative charge, which normally keeps them apart, and their spin, which can be thought of as a tiny bar magnet, pointing either up or down. If electrons of opposite spin overcome their mutual repulsion they can form Cooper pairs and flow without resistance — the essence of superconductivity. One useful way of exploring the properties of a superconductor is to alter this pairing coherence.

Applying a magnetic field disrupts the pairing of electrons, and as a result, diminishes the energy required to break Cooper pairs, thereby reducing the 'transition temperature' at which the material becomes superconducting.

A sufficiently large magnetic field can even disrupt pairing to the point where the material ceases to be superconducting. Adding a single magnetic atom to a superconductor creates a similar effect, although the perturbation to the superconductor is localized near the magnetic atom. On page 920 of this issue, Hudson *et al.*¹ describe the effect of adding a magnetic nickel atom to a high-temperature superconductor. Nickel weakly perturbs its local environment in the superconductor, suggesting that this impurity could be used as a non-invasive probe of superconducting behaviour on the nanoscale.

In a superconductor, pairing coherence occurs when a free electron 'grabs' an electron of opposite spin, forming a Cooper pair and leaving behind a positively charged 'hole' in the electron sea of the superconductor.

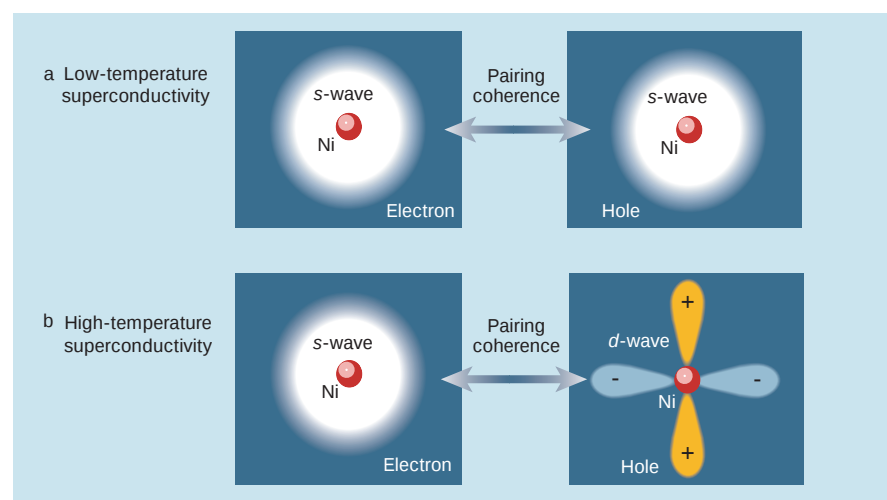


Figure 1 Quasiparticles in a superconductor. An electron bound to an impurity atom, such as nickel (Ni), mixes with a hole, to form a low-energy Bogoliubov quasiparticle. a, In a conventional superconductor, like niobium, the symmetry of the wavefunction of the electron component and of the hole component are identical. They both have s-wave (spherical) symmetry. b, In a high-temperature superconductor, like the one studied by Hudson *et al.*¹, the electrons that form the Cooper pairs rapidly revolve around each other, to keep their distance and lower their mutual repulsion. The d-wave nature of their angular momentum is reflected in the d-wave symmetry of the hole component of the associated quasiparticle.

tor. This mixture of the original electron and the hole left behind is known as a Bogoliubov quasiparticle. The nature of the Cooper pairs, and hence of Bogoliubov quasiparticles, is greatly different in high-temperature superconductors and conventional superconductors. In high-temperature superconductors the two electrons in a Cooper pair revolve around each other rapidly to increase their separation. This reduces the energy cost of having two negatively charged particles close together. Mutual revolution of electrons in a Cooper pair is labelled by angular momentum as s , p or d , just like electrons revolving around the nucleus of an atom. The rapid revolution of the electron pair also causes the Bogoliubov quasiparticles to have a peculiar shape. If the electron component of the quasiparticle is bound to a nickel atom, the hole component must rapidly revolve around the nickel atom, in a way that reflects the nature of the Cooper pairs² (Fig. 1). So in a low-temperature superconductor, the hole would have s -wave symmetry, whereas in a high-temperature superconductor it would have d -wave symmetry.

The nickel atom embedded in the high-temperature superconductor studied by Hudson *et al.*¹ does appear to attract electrons. The authors viewed the area around the nickel atom using a scanning tunnelling microscope (STM), which is sensitive to the density of electronic states near the impurity. They observe two new distinct quasiparticle energies near the nickel atom, corresponding to Bogoliubov quasiparticles with opposite spin. So the experiment of Hudson *et al.* is a direct observation of quasiparticles of distinct spin formed near an individual impurity atom. They have also separately observed the electron and hole components of each quasiparticle by changing the STM voltage from positive (in which an electron is inserted into the superconductor) to negative (in which a hole is inserted). The results directly confirm for the first time that the hole component has the expected d -wave symmetry of a high-temperature superconductor.

Hudson and colleagues' results suggest that the influence of a magnetic nickel atom on its local environment in a high-temperature superconductor is surprisingly weak, despite its strong effect on global properties, such as the superconducting transition temperature. For a conventional superconductor like niobium, embedded magnetic atoms also strongly suppress the transition temperature, and have a weak local effect when placed on the surface³. Other impurity atoms in high-temperature superconductors, such as non-magnetic zinc, have a similar effect to nickel on the transition temperature. But earlier measurements of quasiparticles near a zinc atom⁴ showed dramatic distortions of the local environment. The zinc-associated quasiparticles also lacked any electron component. These differ-

ences suggest that near a zinc atom there may be additional electronic interactions that strongly reduce pairing coherence.

Many studies have shown that high-temperature superconductors doped with nickel behave differently from those doped with other atoms. But Hudson and colleagues' measurements show for the first time that impurity atoms can be present in high-temperature superconductors without dramatically destroying the superconductivity in their local surroundings. The authors suggest that their results have implications for understanding the pairing mechanism underlying high-temperature superconductivity — in particular, whether it has a magnetic origin. It is still too early to favour one mechanism over another, but the importance of their measurements goes beyond these immediate issues.

The focus of high-temperature superconductivity is shifting to the nanoscale, particularly to the effect of small structures in the superconductor on the formation of Cooper pairs. The properties of nickel make it an ideal tool for exploring nanoscale superconductivity. In medical technology, radioactive markers are attached to a molecule to track its motion; here, nickel atoms can be used as 'pairing markers' for inhomogeneous environments in a superconductor. If it is believed that the presence of a particular object in the superconductor, such as a zinc atom or a magnetic field line, destroys pairing coherence over a distance much longer than the atomic spacing, then nickel atoms could be

'seeded' in the vicinity of that object. Then, using STM techniques, any low-energy quasiparticles created near the nickel atom could be monitored for unusual behaviour.

Nickel atoms placed in regions where there is no pairing coherence would display only the electron component of their associated quasiparticles. The hole-like component, which appears only through pairing coherence, would not be visible. One can even imagine circumstances where ordinary quasiparticles found in metals, which are usually pure electron or pure hole, would lose their own coherence. For example, in the case of spin-charge separation the electron itself dissolves into two exotic particles, the 'spinon' and the 'holon'. In this event it is likely that neither an electron nor a hole component would be associated with the nickel atom.

Hudson *et al.*¹ provide a stunning demonstration of the effects of short-range disturbances in high-temperature superconductors. As methods for doping superconductors and modifying their surfaces improve, the possibilities for poking and prodding superconductors on the nanoscale grow even greater.

Michael E. Flatté is in the Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA.

e-mail: michael_flatte@mailaps.org

1. Hudson, E. W. *et al.* *Nature* **411**, 920–924 (2001).
2. Flatté, M. E. & Byers, J. M. *Solid State Phys.* **52**, 137–228 (1999).
3. Yazdani, A., Jones, B. A., Lutz, C. P., Crommie, M. F. & Eigler, D. M. *Science* **275**, 1767–1770 (1997).
4. Pan, S. H. *et al.* *Nature* **403**, 746–750 (2000).

Neuroscience

Awareness of space

Michael S. A. Graziano

Damage to particular parts of the brain can cause spatial confusion and even eliminate awareness of areas of space around the body. The brain regions responsible for spatial awareness, however, are still under debate.

Certain types of brain damage can lead people to suffer from profound spatial disorientation and an inability to use spatial information to direct their movements — symptoms first described in detail near the beginning of the last century^{1,2}. Some of the people who suffer from such damage, to the right posterior parietal lobe of the brain's cortex, show a phenomenon called hemispatial neglect³, whereby they lose their awareness of the half of the space around them that lies opposite to the brain lesion. Someone with hemispatial neglect might shave only half of his face, dress just half of his body, or, when copying a picture, draw only one side of it.

On page 950 of this issue, Karnath and colleagues⁴ revisit the topic of spatial neglect. They look at a set of patients with an

especially pure form of the deficit, who lack some of the other symptoms that can occur together with hemispatial neglect after damage to the posterior part of the brain. The results threaten to overturn the century-old view that the parietal lobe (Fig. 1) is the region of the brain associated with spatial awareness.

The hypothesis that the parietal lobe is the key brain region for processing spatial information is based almost exclusively on studies of humans who have suffered damage to this part of the brain, often as the result of a stroke. Attempts to reproduce hemispatial neglect in monkeys by surgically damaging the parietal lobe have not been successful^{5,6}. Such animals do have defects in processing somatosensory (tactile or heat) stimuli and in directing hand and eye move-

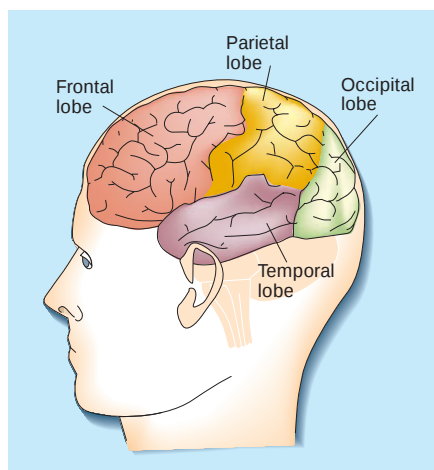


Figure 1 The human brain. Karnath and colleagues⁴ have shown that the topmost part of the temporal lobe is involved in spatial awareness.

ments in space, as do humans with parietal brain damage. But monkeys with parietal damage do not appear to develop hemispatial neglect.

One early study⁷ of the properties of single neurons in the monkey parietal lobe led to the conclusion that much of this region has motor functions — that is, it is involved in producing movements. More recent single-neuron experiments indicated that different subregions of the monkey parietal lobe may be specialized for using sensory information to guide different types of movement, such as movements of the eyes or arms, or shaping the hand for grasping^{8,9}. These subregions connect to specific parts of the motor cortex in the frontal lobe. It has been argued¹⁰ that at least some humans with parietal damage have difficulties mainly in using spatial or sensory information to guide movements, rather than in spatial awareness. So, there is quite strong evidence that the parietal lobe matches sensory input to specific patterns of motor output in both monkeys and humans. But the evidence for a role of the parietal cortex in spatial awareness is less convincing. The classical finding that parietal damage causes hemispatial neglect in humans was ripe for re-examination, and this is what Karnath *et al.*⁴ have done.

The problem is that strokes, or other types of brain damage, usually affect a large part of the brain, so the area responsible for hemispatial neglect in humans has never been known precisely. Karnath *et al.* set out to track down this region. They first selected patients who had damage to the cortex in the right posterior hemisphere. They studied four categories of patients: those with symptoms of hemispatial neglect combined with blindness in parts of the visual field (a typical combination); those with hemispatial neglect but no evidence of blindness; those with blindness but no

hemispatial neglect; and those with neither set of symptoms.

The authors found that the patients with both hemispatial neglect and blindness in parts of the visual field tended to have damage centred on the posterior parietal lobe, fitting with the past century of observations. But the lesions were large, and typically also included the superior temporal lobe. The patients with hemispatial neglect but no blindness had smaller lesions centred on the superior temporal lobe. The few patients with visual-field defects but no neglect had damage that involved the posterior parietal lobe but not the superior temporal lobe. Finally, patients with neither set of symptoms showed no tendency for either the posterior parietal lobe or the superior temporal lobe to be affected. The implication is that damage to the superior temporal lobe is responsible for hemispatial neglect, whereas parietal damage underlies the visual-field defects, perhaps because of injury to the adjacent band of nerve fibres that transmits information to the primary visual cortex. (One might also expect the patients with parietal damage to have problems in integrating sensory and motor information, but this was not tested.)

If confirmed, the finding that spatial neglect is associated with the superior temporal lobe would help to resolve the confusion in the literature. One could then view the human parietal lobe as being involved primarily in matching sensory input and motor output, as it seems to be in monkeys. Spatial awareness or spatial cognition, independent of the control of specific

movements, would depend on other brain structures. These probably include not only the superior temporal lobe but also parts of the frontal lobe, as frontal lesions in both monkeys and humans sometimes cause symptoms of spatial neglect¹¹.

Of course, the assignment of particular functions to specific brain regions can only be taken so far. Brain areas that link sensory and motor information, like those in the parietal lobe, must have some role in spatial awareness, given the apparently close link between planning to make a movement to a location in space and directing attention to that location¹². After all, researchers in the field of systems neuroscience aim not only to determine the properties of specific brain areas, but also to understand how mental functions can emerge from the interactions among different regions.

Michael S. A. Graziano is in the Department of Psychology, Green Hall, Princeton University, Princeton, New Jersey 08544-1010, USA.

e-mail: graziano@princeton.edu

1. Holmes, G. *Br. J. Ophthalmol.* **2**, 449–516 (1918).
2. Balint, R. *Monatsschr. Psychiatr. Neurol.* **25**, 51–81 (1909).
3. De Renzi, E. *Disorders of Space Exploration and Cognition* (Wiley, New York, 1982).
4. Karnath, H.-O., Ferber, S. & Himmelbach, M. *Nature* **411**, 950–953 (2001).
5. Ettlinger, G. & Kalsbeek, J. E. *J. Neurol. Neurosurg. Psychiatry* **25**, 256–268 (1962).
6. Lamotte, R. H. & Acuna, C. *Brain Res.* **139**, 309–326 (1978).
7. Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. *J. Neurophys.* **38**, 871–908 (1975).
8. Snyder, L. H., Batista, A. P. & Andersen, R. A. *Nature* **386**, 167–170 (1997).
9. Sakata, H. & Taira, M. *Curr. Opin. Neurobiol.* **4**, 847–856 (1994).
10. Goodale, M. A. *et al. Curr. Biol.* **4**, 604–610 (1994).
11. Husain, M. & Kennard, C. *J. Neurol.* **243**, 652–657 (1996).
12. Moore, T. & Fallah, M. *Proc. Natl Acad. Sci. USA* **98**, 1273–1276 (2001).

Ecology

Dairy declines hard to swallow

Peter D. Moore

Birds are important constituents of any ecosystem. But they are also valuable monitors of environmental conditions — if their populations begin to decline, we need to know why.

There is no doubt that some populations of farmland birds are in decline in both Europe and North America¹. But what is the cause of such declines? This question cannot usually be tackled experimentally, but writing in the *Journal of Applied Ecology* Anders Møller² describes an instance in which he has been able to do so. The bird concerned is the barn swallow (*Hirundo rustica*). The place is Denmark, where changes in land use have allowed assessment of the effects of declines in dairy farming on swallow populations. Møller finds that the breeding success of swallows depends strongly on the presence of cattle.

Most investigations into population changes in farmland birds are based on cor-

relations, usually relating observed changes in bird density to factors such as the intensity of farming. Correlations are then explained by reference to other factors, such as the availability of insects or seeds as food, resulting in the development of models in which cause and effect can be unravelled. But such reasoning from correlation can all too easily lead the unwary into error.

Møller's approach avoids such pitfalls. He has been able to conduct paired statistical testing of a hypothesis by comparing farms in which land-use changes have taken place with unchanged control farms. He has also been able to take a detailed look at differences in the physical condition and breeding success of his test organism, the barn



Figure 1 In decline — the barn swallow, *Hirundo rustica*. Møller² shows that, on farms where dairy farming has been abandoned, nestlings are smaller and immunologically weaker than on control farms.

swallow, under the two sets of conditions.

Barn swallows are still abundant in both the Old and New Worlds. Their population in Europe has fallen markedly in recent decades, however, possibly by as much as half since 1975. Observations that swallows seem to be more abundant on dairy farms led Møller to the view that a decline in dairy farming could be responsible. To test this hypothesis, he selected 15 farms in which dairy farming had been abandoned and compared them with 15 control farms in which all other factors, such as the availability of nesting habitats, remained unchanged. He checked for alterations in insect food abundance by regular sweep netting, and recorded the number of swallow broods and clutch size, as well as the physical characteristics and growth of nestlings and ringed adult birds, which were regularly netted and then returned to the wild.

Farms where dairy farming had ceased lost around 50% of their insects, and showed an average fall in swallow populations of 48%. The likely reason for this is the loss of tussocky grassland when arable farming is introduced, along with the loss of the dung associated with cattle. The direct cause of the swallow decline was the failure of one-year-old birds to return to the site; recruitment had crashed. The adult birds on both sets of farms did not show any significant changes in body mass, parasite load or any other of the many variables measured, so the problem would appear to lie with the production and survival of young. Nestlings examined on farms where dairy farming had been abandoned were of notably poorer quality, in terms of both size and immune response, than those on control sites. Clutch sizes were also smaller in the absence of cattle. These declines could be detected over the period of change in land use, while unchanged farms showed no such alterations in nestling number or quality.

This thorough and elegant piece of work establishes a clear connection between the presence of dairy cattle and the maintenance of high swallow abundance, operating through the effect of cattle on the density of insects in grassland. Although adult birds can cope with the change, their reproductive success and subsequent recruitment are badly affected. The implications are evident. If the decline in dairy farming in Europe continues, which is likely in view of the current pattern of European Community subsidies that is intended to reduce surplus milk production, then we can expect further declines in swallow populations. A proposal from the Netherlands³, backed by experimentation, suggests that farmers could be encouraged to conserve their bird populations by financial rewards linked to the number of breeding pairs of certain species supported on their land. Perhaps swallows need help of this kind. Equally, birds are sensitive indicators of environmental conditions. When their populations begin to decline — or in some cases rise — unduly, it is usually a sign of changes in the environment.

Meanwhile, in Britain, where the combined effects of BSE and foot-and-mouth disease have led to severe losses in cattle herds, an unexpected repercussion may well be a fall in swallow populations. Other insectivorous birds, such as yellow wagtails, could also be affected. As in Denmark, here is an unsought experimental set-up which, if adequately monitored, could further our understanding of the complex interactions between land use and bird populations. ■

Peter D. Moore is in the Division of Life Sciences, King's College, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK.
e-mail: peter.moore@kcl.ac.uk

1. Hagenmeijer, W. J. M. & Blair, M. J. *The EBCC Atlas of European Breeding Birds* (Poyser, London, 1997).
2. Møller, A. P. J. *Appl. Ecol.* **38**, 378–389 (2001).
3. Musters, C. J. M., Kruuk, M., de Graaf, H. J. & ter Keurs, W. J. *Conserv. Biol.* **15**, 363–369 (2001).

Daedalus

Stored rail energy

A railway, says Daedalus, is a very special kind of road. The distortion of a locomotive wheel must be very slight — the locomotive force must all go into deforming the rail beneath, pushing the top of the rail backwards into shear. The large reaction induced by this force propels the train. If the whole track could be canted backwards in advance, the train would travel by itself. All the energy would be put into the track beforehand and released by the arrival of the train. The track would have to be released behind the driven wheels to couple its stored power to the train, but it seems possible.

So DREADCO engineers are playing with an experimental track. It stores energy by means of a spring that is laid between the running rails and can be canted backwards for ease of deformation. (Ordinary circular compression springs, in which each turn is canted backwards to reduce its height, are used in many types of seal.) The engineers plan to fill the rest of the track with solar panels, so that ordinary daylight will propel the train. During daylight, the panels generate electricity, and the current pushes one turn locally backwards against its neighbours in front. In a few hours the whole spring should be totally canted backwards, storing about 100 W per square metre, which should be plenty. A distributed ratchet system will hold the spring in its energetic, deformed state, storing the solar energy. When the train comes along, a wheel on the spring engaging its local ratcheting is pushed forwards by the release of the canted spring. It is pushed along as the spring recovers its shape, and drives the train wheels over the fixed track. The track is then ready for the next train.

Daedalus freely admits that such a railway is suitable only for occasional traffic. In particular, it can only propel one train per night. Heavier traffic will require a direct input of electricity to the solar panels, which thus become mere accessory power sources. In any case, the idea of storing energy in the track itself, with no locomotive weight to worry about, seems worth exploring.

But Daedalus fears that he has been forestalled. In the recent British Hatfield rail crash, there was an almost total lack of engineering data and just a shamefaced admission that cracks in the rails were somehow to blame. This suggests to Daedalus that ways of storing energy in the running rails had already been tried, and in this case were overdone.

David Jones

Obituary

Frederick Gillett (1937–2001)

As a graduate student at the University of Minnesota, Fred Gillett selected the new field of infrared astronomy as his area of interest, and it became his life's work. He realized that advances in infrared-detector technology would soon allow astronomers to close the enormous gap in the electromagnetic spectrum between visible and radio wavelengths — the only two types of radiation not blocked by the Earth's atmosphere. Every object in the Universe with a temperature above absolute zero radiates in the infrared, so this part of the spectrum contains a great deal of information. Objects that are invisible to optical telescopes can be seen at infrared wavelengths. And, because infrared radiation penetrates dust, even objects that are hidden from view — such as the centre of our Galaxy — can be seen in the infrared.

In 1965, Gillett came to Tucson, Arizona, so he could use my new infrared detector (a low-temperature bolometer that converts heat into electricity). Together we designed and built the first infrared spectrometer capable of measuring radiation at wavelengths of 8 to 14 micrometres from nearby stars. These wavelengths correspond to one of the few 'windows' in the atmosphere through which infrared radiation can pass. One of the first stars observed with the new instrument, Betelgeuse, revealed a broad emission feature that was ultimately identified as common silicate dust in the star's atmosphere. Mature stars produce this cosmic silicate matter as they approach the end of their lives, and it was already known that the next generation of stars are born in regions filled with interstellar dust. So Gillett started his career by revealing one of nature's most successful recycling projects — without which we wouldn't be here.

Gillett received one of the first doctorates in infrared astronomy and introduced infrared instrumentation to the Kitt Peak National Observatory at Tucson (now part of the National Optical Astronomy Observatory). A suite of photometric and spectroscopic instruments and telescope upgrades emerged, along with influential papers on stars, planets and the Galactic Centre. Gillett also delved into the secretive world of strategic and tactical infrared sensors, learning about vital technology not otherwise available to astronomers.

While vigorously pursuing his ground-based activities throughout the 1970s and



Pioneer of infrared astronomy from the ground and in space

1980s, Gillett joined the team that developed the Infrared Astronomy Satellite (IRAS). By operating far above the atmosphere, space telescopes can avoid many of the problems that beset ground-based infrared detectors. As well as becoming one of NASA's most successful international projects, IRAS pioneered the concept of an observatory in space having all of its vital parts cooled to less than two degrees above absolute zero. The talents of many physicists and engineers were required to accomplish this feat, and Gillett supplied his own brand of insightful ideas and quiet determination to get things done.

Launched in 1983 by NASA, in collaboration with Britain and the Netherlands, the 300-day IRAS mission surveyed and mapped 97% of the sky in four broad wavelength bands extending from 8 to 120 micrometres. Its huge impact on astronomy continues to this day. At the top of its long list of discoveries lies the Vega phenomenon, the first observational proof that nearby stars do possess planetary systems like our own. Vega had been chosen as the primary calibration star against which all others were to be compared. The difficult task of establishing and maintaining the absolute calibration of IRAS was assigned to Gillett and his colleague George Aumann. But Vega was not like most other stars; it emitted a huge excess of far-infrared radiation (25–120 micrometres). Gillett

and Aumann quickly solved their calibration problem by selecting other stars that were more mature and much better behaved.

Gillett then reprogrammed IRAS to make observations of Vega and other similar stars. These young stars are orbited by disks of planetary debris not unlike the warmer, but more mature, tenuous dust cloud around the Sun. This discovery launched intensive searches for more stars with planetary debris systems, and they are joined by an ever-growing number of planets as well, suggesting that there may be life on planets other than Earth.

IRAS was followed by two more infrared space missions, COBE in the United States and ISO in Europe, as well as four ambitious infrared projects for the 1990s. These include the Space Infrared Telescope Facility, due to be launched in 2002 and, like IRAS, cooled with superfluid liquid helium, and SOFIA, a large airborne infrared telescope to be carried into the stratosphere by a converted Boeing 747. On the ground there is Gemini, a pair of 8-metre telescopes (one on Mauna Kea, Hawaii, and the other on Cerro Pachón, Chile) and 2MASS, a nearly completed near-infrared (1 to 2.5 micrometre) all-sky survey designed to complement IRAS.

While working for two years as a visiting senior scientist at NASA headquarters, Gillett guided all four projects safely forward to approval, and lived to see them all nearing completion. 2MASS has already uncovered a new class of low-mass stars that bridge the gap between active stars and brown dwarfs, and Gemini has demonstrated unsurpassed infrared performance in early tests. To ensure that the Gemini telescope in Hawaii would be optimized for infrared astronomy, Gillett assumed the role of international project scientist. Despite his rare blood disease, he celebrated the dedication of the first Gemini telescope and witnessed its initial operation.

Gillett enjoyed hiking, especially backpacking in the Grand Canyon, and had a great passion for cycling. He did his part for local fuel conservation by riding his bike more than 20 miles each day to work. He also entered many races, achieving strong finishes in the gruelling El Tour de Tucson bike race. His last major cycling trip was across the state of Washington. He passed away on 22 April in Seattle, Washington, following a stem-cell transplant.

Frank Low

Frank Low is at the Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA. e-mail: flow@theriver.com

The web of human sexual contacts

Promiscuous individuals are the vulnerable nodes to target in safe-sex campaigns.

Unlike clearly defined 'real-world' networks¹, social networks tend to be subjective to some extent^{2,3} because the perception of what constitutes a social link may differ between individuals. One unambiguous type of connection, however, is sexual contact, and here we analyse the sexual behaviour of a random sample of individuals⁴ to reveal the mathematical features of a sexual-contact network. We find that the cumulative distribution of the number of different sexual partners in one year decays as a scale-free power law that has a similar exponent for males and females. The scale-free nature of the web of human sexual contacts indicates that strategic safe-sex campaigns are likely to be the most efficient way to prevent the spread of sexually transmitted diseases.

Many real-world networks¹ typify the 'small-world' phenomenon⁵, so called because of the surprisingly small average path lengths between nodes^{6,7} in the presence of a large degree of clustering^{3,6} (Fig. 1). Small-world networks are classed as single-scale, broad-scale or scale-free, depending on their connectivity distribution, $P(k)$, where k is the number of links connected to a node⁸. Scale-free networks, which are characterized by a power-law decay of the cumulative distribution $P(k) \approx k^{-\alpha}$, may be formed as a result of preferential attachment of new links between highly connected nodes^{9,10}.

We analysed the data gathered in a 1996 Swedish survey of sexual behaviour⁴. The survey involved a random sample of 4,781 Swedes (aged 18–74 years) and used structured personal interviews and question-

naires. The response rate was 59%, which corresponds to 2,810 respondents. Two independent analyses of non-response error revealed that elderly people, particularly women, are under-represented in the sample; apart from this skew, the sample is representative in all demographic dimensions.

Connections in the network of sexual contacts appear and disappear as sexual relations are initiated and terminated. To investigate the connectivity of this dynamic network, in which links may be short-lived, we first analysed the number, k , of sex partners over a relatively short time period — the 12 months before the survey. Figure 2a shows the cumulative distribution, $P(k)$, for female and male respondents. The data closely follow a straight line in a double-logarithmic plot, which is consistent with a power-law dependence. Males report a larger number of sexual partners than females¹¹, but both show the same scaling properties.

These results contrast with the exponential or gaussian distributions — for which there is a well-defined scale — found for friendship networks⁸. Plausible explanations for the structure of the sexual-contact network described here include increased skill in acquiring new partners as the number of previous partners grows, varying degrees of attractiveness, and the motivation to have many new partners to sustain self-image. Our results are consistent with the preferential-attachment mechanism of scale-free networks: evidently, in sexual-contact networks, as in other scale-free networks, 'the rich get richer'^{9,10}.

We next analysed the total number of



Figure 1 It's a small world: social networks have small average path lengths between connections and show a large degree of clustering. Painting by Idahlia Stanley.

partners, k_{tot} , in the respondent's life up to the time of the survey. This value is not relevant to the instantaneous structure of the network, but may help to elucidate the mechanisms responsible for the distribution of number of partners. Figure 2b shows the cumulative distribution, $P(k_{\text{tot}})$: for $k_{\text{tot}} > 20$, the data follow a straight line in a double-logarithmic plot, which is consistent with a power-law dependence in the tails of the distribution.

Our most important finding is the scale-free nature of the connectivity of an objectively defined, non-professional social network. This result indicates that the concept of the 'core group' considered in epidemiological studies¹² must be arbitrary, because there is no well-defined threshold or boundary that separates the core group from other individuals (as there would be for a bimodal distribution).

Our results may have epidemiological implications, as epidemics arise and propagate much faster in scale-free networks than in single-scale networks^{6,13}. Also, the measures adopted to contain or stop the propagation of diseases in a network need to be radically different for scale-free networks. Single-scale networks are not susceptible to attack at even the most connected nodes, whereas scale-free networks are resilient to random failure but are highly susceptible to destruction of the

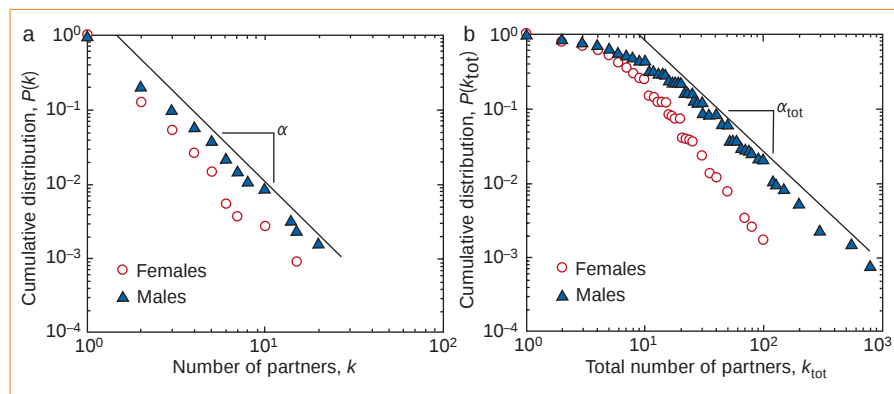


Figure 2 Scale-free distribution of the number of sexual partners for females and males. **a**, Distribution of number of partners, k , in the previous 12 months. Note the larger average number of partners for male respondents: this difference may be due to 'measurement bias' — social expectations may lead males to inflate their reported number of sexual partners. Note that the distributions are both linear, indicating scale-free power-law behaviour. Moreover, the two curves are roughly parallel, indicating similar scaling exponents. For females, $\alpha = 2.54 \pm 0.2$ in the range $k > 4$, and for males, $\alpha = 2.31 \pm 0.2$ in the range $k > 5$. **b**, Distribution of the total number of partners k_{tot} over respondents' entire lifetimes. For females, $\alpha_{\text{tot}} = 2.1 \pm 0.3$ in the range $k_{\text{tot}} > 20$, and for males, $\alpha_{\text{tot}} = 1.6 \pm 0.3$ in the range $20 < k_{\text{tot}} < 400$. Estimates for females and males agree within statistical uncertainty.

best-connected nodes¹⁴. The possibility that the web of sexual contacts has a scale-free structure indicates that strategic targeting of safe-sex education campaigns to those individuals with a large number of partners may significantly reduce the propagation of sexually transmitted diseases.

Fredrik Liljeros*, Christofer R. Edling*, Luís A. Nunes Amaral†, H. Eugene Stanley†, Yvonne Åberg*

*Department of Sociology, Stockholm University, S-106 91 Stockholm, Sweden

e-mail: liljeros@sociology.su.se

†Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

1. Strogatz, S. H. *Nature* **410**, 268–276 (2001).
2. Kochen, M. (ed.) *The Small World* (Ablex, Norwood,

New Jersey, 1989).

3. Wasserman, S. & Faust, K. *Social Network Analysis* (Cambridge Univ. Press, Cambridge, 1994).
4. Lewin, B. (ed.) *Sex i Sverige. Om Sexuallivet i Sverige 1996* [Sex in Sweden. On the Sexual Life in Sweden 1996] (Nat Inst. Pub. Health, Stockholm, 1998).
5. Milgram, S. *Psychol. Today* **2**, 60–67 (1967).
6. Watts, D. J. & Strogatz, S. H. *Nature* **393**, 440–442 (1998).
7. Barthélemy, M. & Amaral, L. A. N. *Phys. Rev. Lett.* **82**, 3180–3183 (1999).
8. Amaral, L. A. N., Scala, A., Barthélemy, M. & Stanley, H. E. *Proc. Natl Acad. Sci. USA* **97**, 11149–11152 (2000).
9. Simon, H. A. *Biometrika* **42**, 425–440 (1955).
10. Barabási, A.-L. & Albert, R. *Science* **286**, 509–512 (1999).
11. Laumann, E. O., Gagnon, J. H., Michael, R. T. & Michaels, S. *The Social Organization of Sexuality* (Univ. Chicago Press, Chicago, 1994).
12. Hethcote, H. W. & Yorke, J. A. *Gonorrhea Transmission Dynamics and Control* (Springer, Berlin, 1984).
13. Pastor-Satorras, R. & Vespignani, A. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
14. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **406**, 378–382 (2000).

Biomechanics

Turning the key on *Drosophila* audition

The genetic dissection of *Drosophila* audition is advancing considerably^{1–4}, but complementary studies are needed to unravel the chain of mechanosensory events that bring about hearing in the fly³. Here we investigate the delicate biomechanics of the fly's minute antennal hearing organs and show that they look and work like a lock and key. Rotating in response to sound, the antenna's distal segment mechanically activates the auditory receptors — thereby 'unlocking' the mechanism

of *Drosophila* audition.

In *Drosophila melanogaster*, the antennae mediate the detection of conspecific 'love songs'^{5,6}. Anatomically, each antenna is an asymmetric structure consisting of three segments and a feather-like arista³ (Fig. 1a). Laser vibrometric analysis of sound-induced vibrations⁷ reveals that the arista and the club-shaped third segment together constitute a mechanical entity — the sound receiver. Both antennal parts oscillate sympathetically in response to acoustic stimulation, consistently exhibiting a moderately damped resonance at 426 ± 16 Hz (quality factor, $Q = 1.2 \pm 0.1$; 8 flies; Fig. 1b).

Mechanical measurements at different locations (Fig. 1b) show that the entire

arista oscillates as a stiff rod, with its vibration velocity continuously decreasing from tip to base. Remarkably, this oscillation is accompanied by rotation of the third segment. The 180° phase shift between the mechanical responses of opposite edges (Fig. 1b) unambiguously shows that they move in opposite directions and that this segment rotates about its longitudinal axis. This rotation is a direct consequence of the radial orientation of the arista, which breaks mechanical symmetry. Physically, the arista introduces a moment arm, increases the effective surface area, and thus determines the twisting force (torque) exerted by sound.

When stimulated acoustically, the third segment rotates like the bow of a key (Fig. 1c). Moreover, the proximal part of this segment presents a stalk that fits into a pit in the second segment, like a key in its lock. Extending along the rotational axis, this stalk transmits the rotation 'downstream', like a key's stem (Fig. 1a, c). Before connecting to the second segment, the stalk bends, forming a hook. This hook then oscillates like a key's bit (Fig. 1a, c). This vibration will maximally stretch and compress the auditory receptors which, notably, are attached perpendicularly to both sides of the hook.

Unlike other animals, *Drosophila* makes use of sound-induced rotation to channel acoustic energy to its auditory receptor neurons — it has 'rotational ears'. This unconventional mechanism relies on a simple trick: the hook balances the asymmetry introduced by the arista, thereby guaranteeing a receptor activation that compares to that in other insect auditory systems⁸. The third segment of the *Drosophila* antenna is also the primary organ of olfaction and carries hundreds of olfactory sensilla⁹. As hearing relies on rotation, the two sensory modalities can co-exist without compromising their respective functions. Such an evolutionary solution is elegant; flies did not turn their nose into an ear, they turn their nose to hear.

Martin C. Göpfert, Daniel Robert

Institute of Zoology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

e-mail: mgopfert@zool.unizh.ch

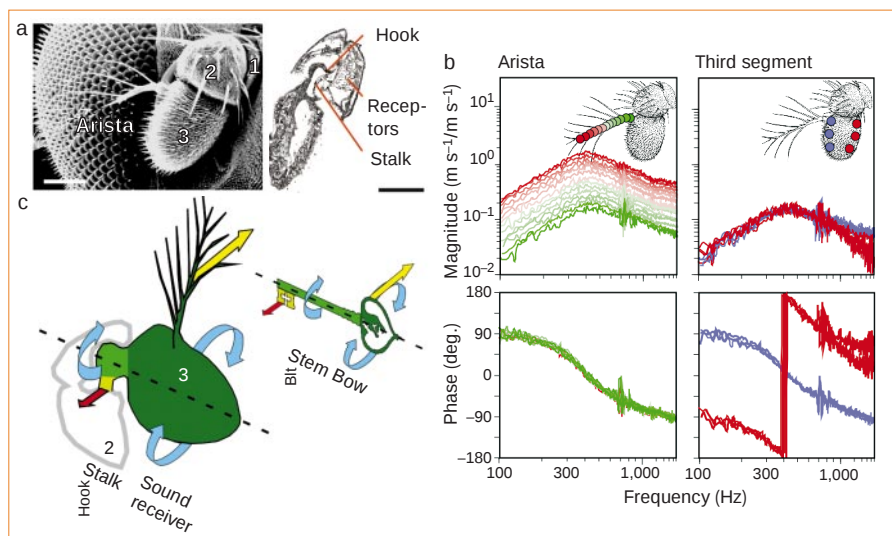


Figure 1 Anatomy, mechanical response and operational mode of *Drosophila* antennal hearing organs. **a**, Anatomy. Left panel, scanning electron micrograph showing the three antennal segments (numbered) and the arista; right panel, longitudinal section through segments 2 and 3. Scale bars, 100 μ m. **b**, Mechanical response to acoustic random-noise stimulation. The arista and the third segment exhibit identical resonance characteristics and rotate about the longitudinal axis of the latter. Frequency spectra show the magnitude and phase of the vibration velocity (measured by microscanning laser Doppler vibrometry⁷) normalized to the particle velocity in the sound field (measured by a particle-velocity microphone at the antenna's position⁷). A response magnitude of unity indicates equal vibration velocity and particle velocity, and a phase of +90° means that the vibration velocity leads the particle velocity by one-quarter of an oscillation cycle. The acoustic stimulus (mean particle velocity ± 0.04 mm s⁻¹) induced maximal displacements of ± 20 nm and rotation of $\pm 0.003^\circ$. Insets, measurement sites and colour convention (drawing by P. Bryant¹⁰). **c**, Operational mode, showing analogy between antennal mechanics and a rotating key. Yellow arrows, input force; blue arrows, rotation; red arrows, output force.

1. Kernan, M., Cowan, D. & Zuker C. *Neuron* **12**, 1195–1206 (1994).
2. Eberl, D. F., Duyk, G. M. & Perrimon, N. *Proc. Natl Acad. Sci. USA* **94**, 14837–14842 (1997).
3. Eberl, D. F., Hardy, R. W. & Kernan, M. J. *J. Neurosci.* **20**, 5981–5988 (2000).
4. Walker, R., Willingham, A. & Zuker C. *Science* **287**, 2229–2234 (2000).
5. Bennet-Clark, H. C. *Nature* **234**, 255–259 (1971).
6. Hall, J. C. *Science* **264**, 1702–1714 (1994).
7. Göpfert, M. C. & Robert, D. *Proc. R. Soc. Lond. B* **267**, 453–457 (2000).
8. Yager, D. D. *Microsc. Res. Tech.* **47**, 380–400 (1999).
9. Carlson, J. R. *Trends Genet.* **12**, 175–180 (1996).
10. FlyBase consortium (<http://flybase.bio.indiana.edu>) *Nucleic Acids Res.* **27**, 85–88 (1999).

Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution

Patrick Jordan*, Petra Fromme†, Horst Tobias Witt†, Olaf Klukas*, Wolfram Saenger* & Norbert Krauß*‡

* Institut für Chemie/Kristallographie, Freie Universität Berlin, D-14195 Berlin, Takustraße 6, Germany

† Max Volmer Laboratorium für Biophysikalische Chemie, Institut für Chemie, Fakultät 2; Technische Universität Berlin, D-10623 Berlin, Straße des 17. Juni 135, Germany

Life on Earth depends on photosynthesis, the conversion of light energy from the Sun to chemical energy. In plants, green algae and cyanobacteria, this process is driven by the cooperation of two large protein–cofactor complexes, photosystems I and II, which are located in the thylakoid photosynthetic membranes. The crystal structure of photosystem I from the thermophilic cyanobacterium *Synechococcus elongatus* described here provides a picture at atomic detail of 12 protein subunits and 127 cofactors comprising 96 chlorophylls, 2 phylloquinones, 3 Fe_4S_4 clusters, 22 carotenoids, 4 lipids, a putative Ca^{2+} ion and 201 water molecules. The structural information on the proteins and cofactors and their interactions provides a basis for understanding how the high efficiency of photosystem I in light capturing and electron transfer is achieved.

All higher organisms on earth receive their energy directly or indirectly from oxygenic photosynthesis performed by plants, green algae and cyanobacteria, the last two representing an important part of the marine food web. Photosynthesis converts solar energy to chemical energy by means of two large protein complexes located in the thylakoid membranes: photosystems I (PSI) and II (PSII). They use the energy of absorbed photons to translocate electrons across the membrane along a chain of cofactors, leading to a transmembrane difference in the electrical potential and in H^+ concentration, which drives ATP synthesis and reduction of NADP^+ to NADPH. In the subsequent dark reactions, NADPH and ATP are used to convert CO_2 to carbohydrates.

PSI and PSII belong to the superfamily of photosynthetic reaction centres, which are divided into two classes according to their terminal electron acceptors: Fe_4S_4 clusters (type I) or quinone (type II; for review see ref. 1). For purple bacterial reaction centres and PSII, both belonging to the type II reaction centres, three-dimensional structures are known at high and medium resolution, respectively^{2,3}. For type I reaction centres, a structural model of cyanobacterial PSI at medium resolution (4 Å) is available^{4,5} but elucidation of the mechanism of light capture and electron transfer in PSI has been impeded by lack of more detailed structural information.

Isolated cyanobacterial PSI exists as a trimer⁶ (relative molecular mass M_r 3 × 356,000); this form is also present *in vivo*. One monomer consists of at least 11 different protein subunits coordinating more than 100 cofactors. PSI captures light energy by a large internal antenna system and guides it to the core of the reaction centre with high efficiency. After primary charge separation initiated by excitation of the chlorophyll dimer P700, the electron passes along the electron transfer chain (ETC) consisting of the spectroscopically identified cofactors A_0 (Chl a), A_1 (phylloquinone) and the Fe_4S_4 clusters F_X , F_A and F_B (for review see ref. 7). At the stromal (cytoplasmic) side, the electron is donated by F_B to ferredoxin (or flavodoxin) and thence transferred to NADP^+ reductase. The reaction cycle is completed by re-reduction of P700^{+*} by cytochrome c_6 (or plastocyanin) at the inner (luminal) side of the membrane. The electron carried by cytochrome c_6 is

provided by PSII by way of a pool of plastoquinones and the cytochrome b_6/f complex.

To understand the role of the protein subunits in binding of the cofactors, and the interactions between the cofactors, we have determined the X-ray crystal structure of PSI from the thermophilic cyanobacterium *S. elongatus* and describe here the atomic model.

Structure determination

The trimeric PSI, isolated from *S. elongatus*, was crystallized as described elsewhere⁸. Detailed information on X-ray data collection, evaluation and refinement at 2.5 Å resolution are provided in Table 1 and as Supplementary Information.

In the electron density map, the 11 biochemically characterized and sequenced protein subunits⁹ were identified. Ten of these were interpreted with the known amino-acid sequences, whereas the peripheral subunit PsaK could only be modelled as polyalanine, because the electron density becomes less well defined with increasing distance from the centre of the PSI trimer. A twelfth polypeptide chain containing one transmembrane α -helix could be fitted to well defined electron density. Its sequence is homologous to the controversial subunit PsaX identified in PSI from the thermophilic cyanobacteria *Synechococcus vulcanus* and *Anabaena variabilis*^{10,11}. For the 96 chlorophylls, the positions and orientations of the chlorophyll head groups were determined, leading to the assignment of the orientation of the Q_X and Q_Y transition dipolar moments. For 48 of them the ring substituents, including the complete phytyl side chains, could be modelled into the electron density map, whereas for the remaining Chl a molecules the phytyl chain could be modelled only partially. In addition, the three Fe_4S_4 clusters, two phylloquinones and for the first time 22 carotenoids (16 in all-*trans*, 5 in different *cis* configurations and one modelled as an incomplete β -carotene molecule), 4 lipids and a putative Ca^{2+} ion were identified. The total of 127 cofactors located in one PSI monomer contributes around 30% of the molecular mass of PSI.

Overall architecture

Trimeric PSI resembles a clover leaf (Fig. 1a). Its threefold rotation axis coincides with the crystallographic C_3 axis (space group $P6_3$), which is perpendicular to the membrane plane (Fig. 1b). At the

‡ Present address: Institut für Biochemie, Universitätsklinikum Charité, Medizinische Fakultät der Humboldt Universität zu Berlin, Monbijoustraße 2, D-10117 Berlin, Germany.

'trimerization domain' close to the C_3 axis, PsaL forms most of the contacts between the monomers.

PSI contains nine protein subunits featuring transmembrane α -helices (PsaA, PsaB, PsaF, PsaI, PsaJ, PsaK, PsaL, PsaM and PsaX) and three stromal subunits (PsaC, PsaD and PsaE). The large (M_r 83K) subunits PsaA and PsaB are related by a pseudo- C_2 axis located at the centre of the PSI monomer and oriented parallel to the C_3

axis. The organic cofactors of the ETC are arranged in two branches along the pseudo- C_2 axis, and most of the antenna Chl a molecules, the carotenoids and the lipids are bound to PsaA/B. The other membrane-intrinsic subunits are peripheral to the PsaA/PsaB core and contribute to the coordination of antenna cofactors (Fig. 1a).

The structure of PSI suggests a docking site for cytochrome c_6 or plastocyanin (donating electrons to PSI) at the luminal side close to

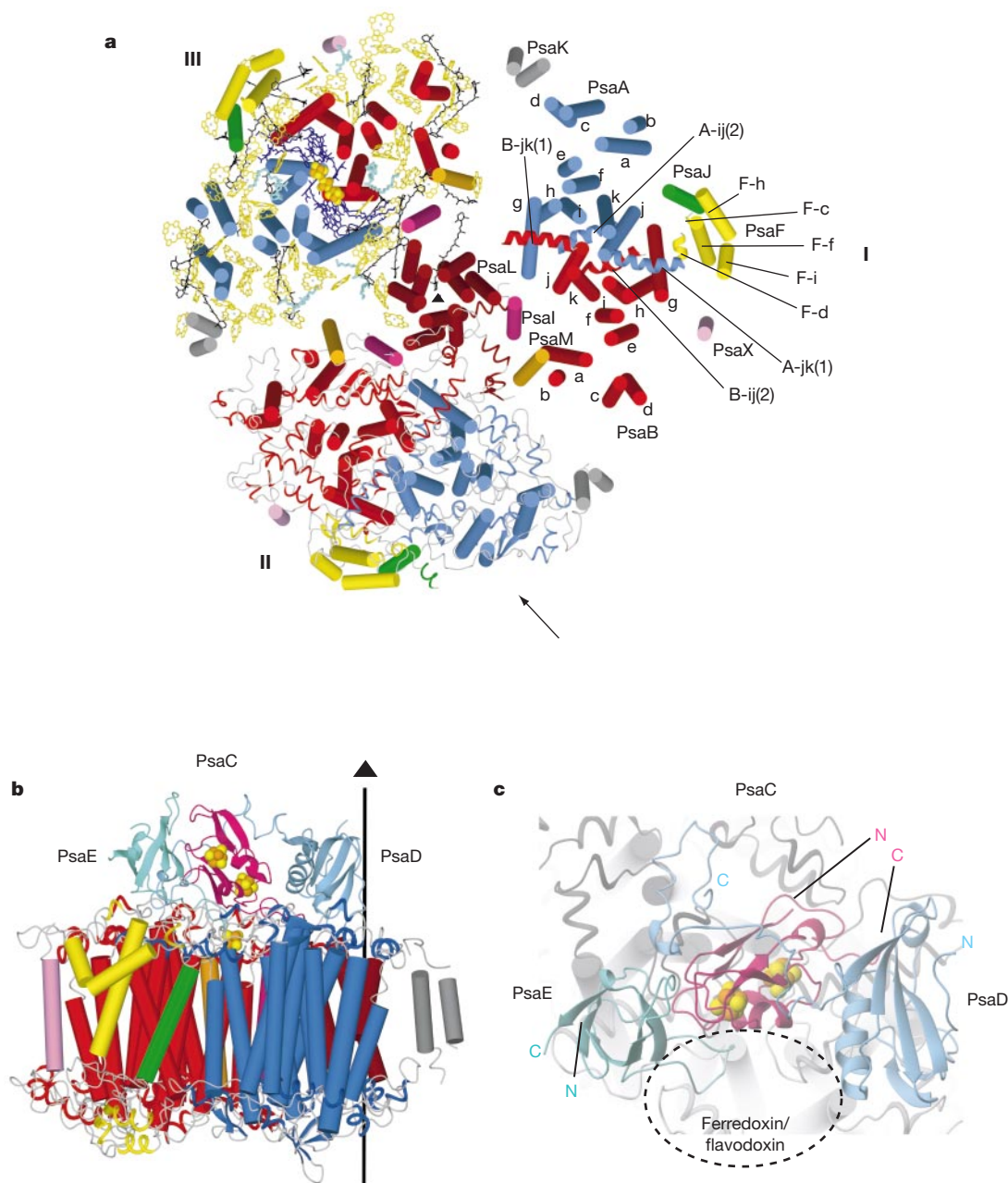


Figure 1 Structural model of PS I trimer at 2.5 Å resolution. **a**, View along the membrane normal from the stromal side. For clarity, stromal subunits have been omitted. Different structural elements are shown in each of the three monomers (I, II and III). I, arrangement of the transmembrane α -helices (cylinders). Subunits are labelled. The transmembrane α -helices of PsaA (blue) and PsaB (red) are named A-a to A-k (B-a to B-k) from the N to the C terminus (capital letters omitted). All loop regions of PsaA and PsaB are named according to the transmembrane helices which they connect. For all other subunits the α -helices and β -sheets are numbered in alphabetical order from the N to the C terminus. Six helices in extra-membranous loop regions are drawn as spirals. II, membrane-intrinsic subunits. In addition to the transmembrane α -helices of the stromal and luminal loop regions are shown in ribbon representation. III, complete set of cofactors shown with the

transmembrane α -helices (the side chains of the antenna Chl a molecules have been omitted). ETC: quinones and chlorophylls in blue, iron and sulphur atoms of the three Fe_4S_4 clusters as orange and yellow spheres, respectively. Antenna system: chlorophylls in yellow, carotenoids in black, lipids in turquoise. **b**, Side view of the arrangement of all proteins in one monomer of PSI (colours as in **a**), including the stromal subunits PsaC (pink), PsaD (turquoise), PsaE (green) and the Fe_4S_4 clusters. View direction indicated by arrow at monomer II in **a**. The vertical line (right) shows the crystallographic C_3 axis. **c**, View as in **a** showing stromal subunits PsaC, PsaD and PsaE. They cover some of the loop regions and helices of PsaA and PsaB (light grey). Dashed ellipse: putative docking site of ferredoxin, covering loops of PsaA.

Table 1 Data collection, phasing and refinement statistics

Data set	Native (1)	Native (2)	EMTS	PIP
Wavelength (Å)	0.99	1.44	0.99	0.99
Resolution (Å)	30–2.5	30–2.9	50–3.0	50–3.0
R_{merge}^*	0.064 (0.247)	0.079 (0.207)	0.076 (0.224)	0.065 (0.254)
Completeness (%)	97.1 (88.9)	90.4 (76.3)	98.8 (95.6)	93.3 (68.3)
$\langle I \rangle$	10.6 (3.4)	10.8 (4.6)	20.5 (8.9)	21.1 (4.2)
Phasing statistics				
No. of heavy atom sites	–	–	4	8
R_{cullis} (centric)†	–	–	0.86	0.81
Phasing power (iso/ano)†	–	–/0.51	1.23/1.03	1.62/1.07
Resolution (Å)	30–2.5	30–3.0	30–3.0	30–3.0
FOM‡	0.57	–	–	–
Refinement statistics				
Reflections (working set)	233,377	–	Number of non hydrogen atoms	
Reflections (test set)	4,743	–	Protein	17,404
$R_{\text{work}}/R_{\text{free}}$ (%)	19.8/21.7	–	Cofactors	6,602
R.m.s.d. bond lengths (Å)	0.0126	–	Water	201
R.m.s.d. bond angles (°)	1.475	–	Metal	1
Coordinate error (Å)‡	–	–	–	–
Luzzati	0.32	–	–	–
SigmaA	0.36	–	–	–

Numbers in parentheses correspond to values in the highest resolution shell.

* $R_{\text{merge}} = \sum_i \sum_h |I(h) - \langle I(h) \rangle| / \sum_i \sum_h I(h)$, where $I(h)$ are individual intensities of any reflection and $\langle I(h) \rangle$ is the mean intensity.

†FOM (figure of merit), R_{cullis} (centric) and phasing power determined by the program SHARP (see Supplementary Information).

‡From Luzzati plot and SigmaA analyses, as determined with CNS (see Supplementary Information).

the pseudo- C_2 axis and for ferredoxin or flavodoxin (accepting electrons from PSI) at the stromal side. The binding pocket for the latter is formed by subunits PsuC, PsuD and PsuE (Fig. 1c).

Polypeptide subunits

Subunits PsuA and PsuB share similarities in protein sequence and structure. They contain 11 transmembrane helices each that are divided into an amino-terminal domain (six α -helices: A/B-a to A/B-f) and a carboxy-terminal domain (five α -helices: A/B-g to A/B-k); for nomenclature see Fig. 1. The latter form two interlocked semicircles enclosing the ETC cofactors. This core structure is separated from the N-terminal α -helices A/Ba-d and the transmembrane α -helices of the smaller PSI subunits by an elliptically distorted cylindrical region bridged by α -helices A/B-e and f and harbouring a large number of the antenna Chl_a molecules and carotenoids (see Fig. 1a and ‘The core antenna system’ section).

The ‘flat but rugged’ luminal surface of PSI is mainly formed by loop regions connecting transmembrane α -helices of subunits PsuA and PsuB. The amino-acid sequences of the loops contain segments not conserved between PsuA and PsuB, breaking the pseudo- C_2 symmetry. The loops, forming several short α -helices and β -sheets, shield the cofactors from the aqueous phase. A hollow close to the pseudo- C_2 axis has been suggested to be the binding site for the electron carriers cytochrome c_6 or plastocyanin¹². The base of the hollow is formed by α -helices A/B-ij(2) of loops A/B-ij in which two conserved Trp residues A655 and B631 are located. Their indole groups are stacked at van der Waals distance and related by the pseudo- C_2 axis. As they point into the putative binding site, they may be engaged in interaction with and/or electron transfer from cytochrome c_6 or plastocyanin to the oxidized primary electron donor P700⁺. This view is supported by mutagenesis studies on cyanobacterial PSI where substitution of some of the amino acids of A/B-ij(2) including Trp affects electron transfer to P700¹³.

Besides the luminal loops of PsuA/B, subunit PsuF contributes prominent structural features to this surface of PSI with two hydrophilic α -helices F-c and F-d at the N terminus of transmembrane helix F-f (Fig. 1a, b). As the shortest distance between their helix axes and the pseudo- C_2 axis is 27 Å, direct interaction with cytochrome c_6 or plastocyanin is unlikely, consistent with the observation that deletion of PsuF does not influence the kinetics of electron transfer in cyanobacteria¹⁴. By contrast, an insertion and an extension in the N-terminal domain in PsuF of higher plants (see ref. 15 and references therein) are involved in plastocyanin docking.

The loops at the stromal surface of PsuA/B partly constitute the binding interface to subunits PsuC, PsuD and PsuE. These subunits are close together and arranged as a crescent, with PsuD closest to the C_3 axis, PsuC in the middle and PsuE farthest out (Fig. 1b, c). They surround an indentation on PsuA that was proposed as the binding site for ferredoxin^{12,16}.

PsuC harbours the two Fe_4S_4 clusters F_A and F_B . It exhibits pseudo-twofold symmetry similar to bacterial $2\text{Fe}_4\text{S}_4$ ferredoxins¹⁷, but contains an insertion of 10 amino acids in the loop connecting the iron-sulphur cluster binding motifs and extensions of the N and C termini by 2 and 14 amino acids, respectively (Fig. 2). As the insertion extrudes as a large loop, it may be engaged in docking of ferredoxin or flavodoxin. The long C terminus of PsuC (green in Fig. 2) interacts with PsuA/B/D and appears to be important for the proper assembly of PsuC into the PSI complex.

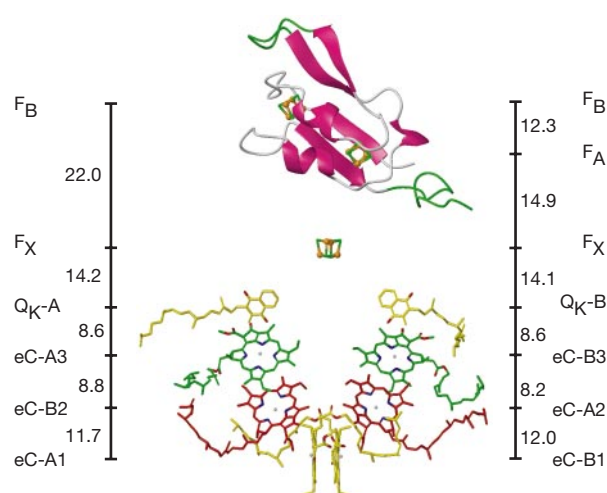


Figure 2 Cofactors of the electron transfer chain (ETC) and of PsuC. View parallel to the membrane plane. The pairs of chlorophylls of the ETC are arranged in two branches A and B. They are labelled eC, followed by the letter A or B indicating whether PsuA or PsuB, respectively, coordinates Mg^{2+} , and by numbers 1 to 3 starting from the luminal side. Phylloquinones are QK-A and QK-B. The Fe_4S_4 clusters are labelled F_X , F_A and F_B according to their spectroscopic terms. The centre-to-centre distances between the cofactors (black lines) are given in Å. In PsuC, parts analogous to $2[\text{Fe}_4\text{S}_4]$ bacterial ferredoxins¹⁷ are pink, an insertion and extensions at C and N termini are green.

Table 2 Coordination of the antenna Chla

Central ligands of the antenna Chla											
	His	Asn	Gln	Asp	Glu	Tyr	Backbone	Phospho-diester	Water	Unknown ligand	Total
All subunits	65		3	1	2	1	1		15	1	90
PsaA	32		2				1		5		40
PsaB	29		1	1		1			7		39
PsaJ	1				1				1		3
PsaK	(1)*									1	2
PsaL	1				1				1		3
PsaM									(1)*		1
PsaX		1									1
PG								1			1

Sectioning of antenna Chla coordinated by Psa/PsaB into peripheral and central antenna domains†

N-terminal sequence region								C-terminal sequence region				
		transm. α -helices				Loops		transm. α -helices			loops	
		a	b	e	f	N-ter†	ab	g	h	k	fg	jk
central domain	PsaA	3	1	1	3	2	2	3	3	1	1	2
	PsaB	3	—	1	3	1	3	3	3	1	1	2
		b	c	d	e	cd	de				gh	
peripheral domain	PsaA	1	4	3	2	1	5	—	—	—	2	—
	PsaB	1	4	3	2	—	5	—	—	—	3	—

* Numbers in parentheses correspond to suspect ligands.

† Secondary structural elements are labelled according to the scheme presented in legend to Fig. 1.

‡ Loops N-terminal of transmembrane α -helices A/B-a.

PsaD forms an antiparallel, four-stranded β -sheet, in which the loop connecting the third and fourth strands contains an α -helix, followed by a two-stranded β -sheet. The loop segment reaching from His D95 to the C terminus is attached by numerous hydrogen bonds to the stromally exposed sides of PsaC and E, and seems to help in the positioning of these subunits, as suggested by the importance of PsaD for electron transfer from F_X to F_A/F_B .

The structure of PsaE consists of a five-stranded antiparallel β -barrel similar to that found in NMR studies on PsaE from *Synechococcus* PCC7002 and *Nostoc* sp.^{18,19}. The long loop connecting β -strands c and d seen here, however, is in a conformation different from that in solution, probably because it interacts with PsaC and the stromal sides of PsaA and PsaB.

Together with PsaA and PsaB, the small subunits featuring transmembrane α -helices bind and stabilize the cofactors of the core antenna system. PsaJ, PsaK, PsaL, PsaM and PsaX coordinate the Mg^{2+} ions of antenna Chla molecules directly or through water molecules (Table 2). Subunits PsaF, PsaI, PsaJ, PsaL, PsaM and to a lesser extent PsaK are in hydrophobic contact with the carotenoids. In the trimerization domain, PsaL contributes to most contacts between monomers, supported by PsaI, PsaB and PsaA. According to the electron density map, the C terminus of PsaL is six amino-acid residues longer than the published sequence⁹ but agrees with the complementary DNA sequence if a frame shift is accounted for. A high spherical electron density (at 5σ ($F_O - F_C$)) located at the interface between adjacent monomers and surrounded by Asp side chains of PsaL and PsaA may belong to a metal ion that could be important for trimer formation. It might be a Ca^{2+} as found in trimeric but not in monomeric PSI from *Synechocystis* PCC6803 (P. Chitinis, personal communication).

The electron transfer chain

This functionally most important part of PSI is formed by six chlorophylls, two phylloquinones and three Fe_4S_4 clusters (Fig. 2). The chlorophylls and phylloquinones are arranged in two branches A and B as pairs of molecules related by the pseudo- C_2 axis and coordinated to PsaA and PsaB. The A branch features chlorophylls eC-A1, eC-B2, eC-A3 and phylloquinone Q_K-A , and the B branch features chlorophylls eC-B1, eC-A2, eC-B3 and phylloquinone Q_K-B (for nomenclature see Fig. 2). The cofactors forming one branch are not bound exclusively to one subunit of PSI, the crossovers being

eC-B2 and eC-A2.

Charge separation is initiated at the primary electron donor P700 formed by the chlorophyll pair eC-A1/eC-B1. The chlorin planes of these chlorophylls are parallel at 3.6 Å interplanar distance and oriented perpendicular to the membrane plane. Rings I and II overlap partially with the Mg^{2+} ions separated by 6.3 Å (Fig. 3a). The structure of P700 differs from the 'special pair' formed by bacteriochlorophylls in PbRC, where the Mg^{2+} separation is larger (7.6 Å)²⁰ than in P700 and π - π interaction can be assumed to be stronger because of the perfect overlap of rings I²¹.

In contrast to the homodimeric special pair in PbRC, P700 is a heterodimer, eC-B1 being Chla (Mg^{2+} coordinated to His B660) and eC-A1 being chlorophyll a' (Chla'), the C13² epimer of Chla (Mg^{2+} coordinated to His A680). This is consistent with chlorophyll extraction experiments suggesting that P700 contains one or two Chla' molecules²². The asymmetry extends to the binding pockets for the two chlorophylls. Whereas the side chains of transmembrane α -helices A-i and A-k and a water molecule form hydrogen bonds to Chla' eC-A1 (Fig. 3a, b), no hydrogen bonds are found to Chla eC-B1. In analogy to ref. 23, the electron spin density in P700⁺ should be localized to a larger extent on the non-hydrogen-bonded Chla (eC-B1), in agreement with mutagenesis²⁴ and electron nuclear double resonance (ENDOR) studies²⁵, showing that $\approx 80\%$ of the spin density in P700⁺ is probably located on the chlorophyll bound to PsaB.

The chlorophylls of the second pair of Chla molecules next to P700, eC-A1/eC-B2, are not engaged in any hydrogen bonds. Their Mg^{2+} ions are coordinated by water molecules hydrogen-bonded to side chains of Asn A604 and Asn B591, respectively (Fig. 4a).

One or both of the Chla molecules of the third pair of chlorophylls eC-A3/eC-B3 probably represents the electron acceptor A_0 . The Chla molecules are coordinated to sulphur atoms of Met A688 and Met B668 (distance $Mg^{2+}-S = 2.6$ Å), respectively (Fig. 4a). This is remarkable, as interactions between the hard acid Mg^{2+} and the soft base S should be weak according to the concept of hard and soft acids and bases. To our knowledge, this coordination of Mg^{2+} is unique in proteins. It remains to be elucidated whether the $Mg^{2+} \cdots S$ coordination could contribute to the unusually low redox potential (approximately -105 mV) of electron acceptor A_0 (ref. 7). Short hydrogen bonds are donated by the hydroxyl groups of tyrosines A696 and B676 ($d_{O-O} = 2.7$ Å) to the keto oxygens of rings V of

eC-A3 and eC-B3, respectively. As the chlorin planes of eC-B2 and eC-A2 are roughly parallel with and only ~ 3.8 Å from those of eC-A3 and eC-B3, respectively, they may be directly engaged in electron transfer from P700 to A_0 and possibly affect the spectroscopic and redox properties of the third Chl_a pair.

All the amino acids involved in Mg^{2+} coordination and hydrogen bonding to the second and third Chl_a pairs of the ETC are strictly conserved between PsaA and PsaB from cyanobacteria to higher plants, suggesting that these protein–chlorophyll interactions are essential for fine-tuning the redox potentials of the cofactors. As the edge-to-edge distances between adjacent chlorophylls of the ETC are very short (~ 4 Å), we expect that each participates in electron transfer.

One or both of the phyloquinones Q_K -A and Q_K -B might correspond to the electron acceptor A_1 . Their quinone planes are π -stacked at 3.0–3.5 Å with indole rings of conserved Trp A697 and Trp B677 on the stromal surface α -helices A/B-jk(1), respectively. This agrees with modelling based on electron paramagnetic resonance (EPR) data²⁶ and quinone reconstitution studies²⁷. Only the

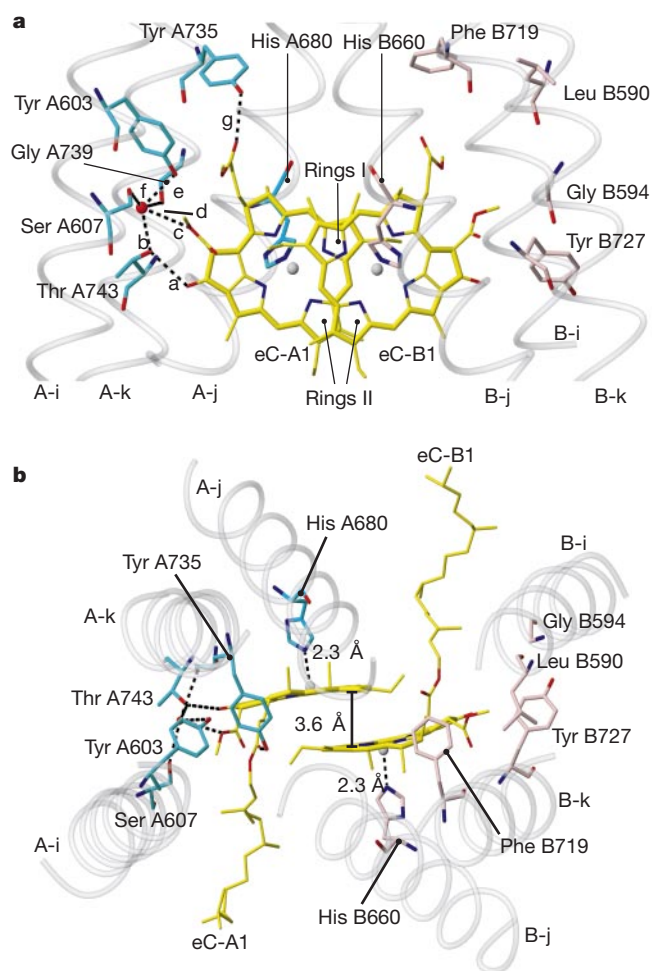
carbonyl oxygen atom in ortho position to the phytol chain accepts hydrogen bonds from backbone NH of Leu A722 (Q_K -A) and Leu B706 (Q_K -B) (Fig. 4c, d). It remains to be seen whether the lack of a hydrogen bond to the para-carbonyl oxygen contributes to the low redox potential (approximately -820 mV) of the phyloquinone in PSI, whereas that of the quinone Q_A in PbcRC and in PSII is around -130 mV.

A controversial issue in PSI is whether one or both branches of the ETC are active (for review see refs 7, 28). Spectroscopic data on $P700^{++}$ and the radical pair $P700^{++}/A_1^{\bullet-}$ suggest an asymmetry of electron transfer along the ETC in PSI. P700 is the only site where a significant structural difference between the two branches is obvious. The spin-carrying half of $P700^{++}$ is probably Chl_a eC-B1, yielding two possible radical pairs: $eC-B1^{++}/Q_K-B^{\bullet-}$ and $eC-B1^{++}/Q_K-A^{\bullet-}$, with centre-to-centre distances of 24.4 ± 0.3 Å and 26.0 ± 0.3 Å, respectively. As the latter value agrees better with the distance of 25.4 ± 0.3 Å between $P700^{++}$ and $A_1^{\bullet-}$ determined by EPR spectroscopy at 80 K²⁹, this phyloquinone might be Q_K -A, as also suggested by mutagenesis studies³⁰. According to kinetic investigations, electron transfer in PSI involves both branches of the ETC with different rate constants^{31,32} of 35×10^6 s⁻¹ and 4.4×10^6 s⁻¹ for the electron transfer steps from each phyloquinone to F_X . Mutation of Trp A697 in the Q_K -A binding site decreases the slower rate, indicating that the A branch is the slower one^{31,32}. Surprisingly, the edge-to-edge distances (which according to the Moser–Dutton approximation³³ determine the optimal electron transfer rates) are identical ($R = 6.8$ Å) between Q_K -A/ Q_K -B and F_X , and would correspond to optimal rates³³ of 8.3×10^{10} s⁻¹ for the corresponding electron transfer steps in both branches, orders of magnitude faster than the values for the experimental rate but in agreement with an estimation of 9×10^{10} s⁻¹ derived from the temperature dependence of the slower experimental rate³⁴. The different transfer rates may be caused by the most obvious elements of asymmetry in the vicinity of the two binding pockets: Trp B673 (not conserved in PsaA) located between Q_K -B and F_X , the negatively charged phospholipid (I) at 15 Å from Q_K -A, contrasting to the uncharged galactolipid (II) at the same distance from Q_K -B and differences in the number and arrangement of water molecules located between Q_K -A, Q_K -B and F_X . Other possible explanations might relate to minor structural differences not detectable at the present resolution.

The pseudo- C_2 axis passes right through the [4Fe4S] cube of F_X (ref. 5), which is a rare example of an inter-polypeptide iron sulphur cluster. F_X is coordinated to both PsaA and PsaB in strictly conserved loop segments A/B-hi containing two cysteines each (Fig. 4b), as proposed from the amino-acid sequences³⁵. A possible geometrical distortion of F_X associated with the unusually low redox potential of around -0.7 mV (ref. 36) is not obvious at the present resolution. The assignment of the two Fe_4S_4 clusters in PsaC to F_A and F_B is now clear, as the cysteine residues of PsaC coordinating F_A and F_B are known from mutational studies³⁷ and are well defined in the structure. The arrangement of the clusters with F_A being closer to F_X than F_B suggests a sequence $F_X \rightarrow F_A \rightarrow F_B$ in electron transfer, in agreement with spectroscopic data (ref. 38 and refs therein).

The core antenna system

In contrast to PbcRC, which collects light energy through separate membrane-intrinsic light-harvesting complexes, PSI features a core antenna system formed by 90 Chl_a molecules and 22 carotenoids (Figs 1 and 5). Seventy-nine of the antenna Chl_a molecules are bound to PsaA and PsaB (Table 2), but the small subunits PsaJ, PsaK, PsaL, PsaM, PsaX and lipid (III), a phosphatidylglycerol, also coordinate to Mg^{2+} of 11 Chl_a molecules either directly or indirectly through water molecules (Table 2). Except for aC-A40 and aC-B39, the 43 innermost Chl_a molecules of the antenna fill the volume provided by the elliptically distorted cylindrical region centred on the pseudo- C_2 axis and are separated from the chlorophylls of the



ETC by not less than ~ 18 Å (Fig. 5). The antenna is extended by 2 sets of 18 peripheral Chl_a molecules bound to PsaA and PsaB, forming layer-like structures at the stromal and luminal sides (Fig. 5b and Table 2). The arrangement of the PsaA/B-bound antenna Chl_a molecules follows the pseudo-C₂ symmetry in a loose sense. Almost all of the Chl_a molecules are at Mg²⁺–Mg²⁺ distances between 7 and 16 Å (Fig. 5c), a range supporting fast excitation energy transfer of the Förster type³⁹. Although most of the Chl_a molecules are coordinated to His imidazoles, some are coordinated to oxygen atoms of Gln, Asp, Glu and Tyr side chains, to peptides or to water oxygens (Table 2).

Two of the Chl_as (aC-A40 and aC-B39) are special as they structurally and perhaps functionally connect eC-A3 and eC-B3 of the ETC to the antenna (Fig. 5c). The centre-to-centre distances are 12.8 Å between aC-A40 and eC-A3 and 10.9 Å between aC-B39 and eC-B3. Otherwise, the ETC and antenna are well separated and isolated from each other. It is not yet clear whether transfer of energy proceeds through these two Chl_as from the antenna to P700. If this were the case, eC-A2, eC-A3, eC-B2 and eC-B3 could be engaged both in excitation energy and in electron transfer.

The 'red' chlorophylls

In *S. elongatus* PSI, 9 to 11 Chl_a molecules per monomer absorb

light at longer wavelengths (bathochromically or 'red-shifted') than P700 (ref. 40), possibly facilitating light energy capture under extreme environmental conditions or focussing excitation energy to the core of the reaction centre (for review see ref. 41). Among several causes of the red shift⁴¹, excitonic coupling of Chl_a can be most safely deduced from the present structure. Putative candidates for strongly excitonically coupled Chl_a molecules are one trimer and three dimers. In the Chl_a trimer at the luminal side of the membrane and close to PsaX, aC-B31, aC-B32 and aC-B33 are stacked like a staircase (Fig. 5), with interplanar separations within 3.5–3.7 Å and lateral centre-to-centre shifts of 8.3 Å. Of these Chl_as, only aC-B31 is directly coordinated to His B470, the Mg²⁺ of the other two being coordinated to water. The pseudo-translational symmetry of this trimer implies that the transition dipole moments of the chlorins are roughly parallel, reminiscent of crystal structures of chlorophyll derivatives⁴², with absorption maxima at 740 nm. This extreme red shift compared to the solution state is thought to derive from strong excitonic coupling supported by π – π interactions within the stacks. The Chl_a trimer is singular in PSI, possibly because a counterpart in PsaA related by the pseudo-C₂ symmetry would interfere with the monomer–monomer contact in the PSI trimer and the hydrophobic segment stabilizing the Chl_a trimer in PsaB is deleted in PsaA.

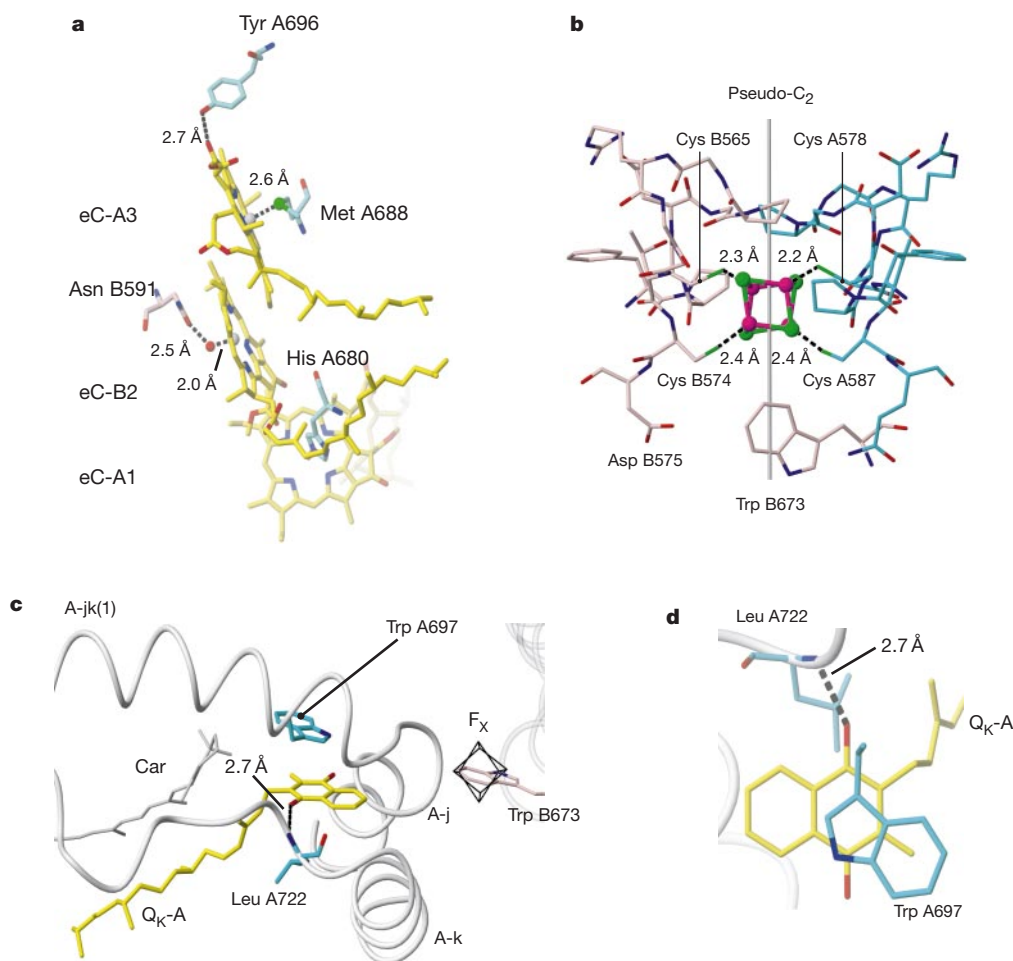


Figure 4 Local environment of membrane-intrinsic cofactors of the electron transfer chain. **a**, Chlorophylls of the A branch of the ETC, with axial ligands and hydrogen bonds involving chlorophylls eC-B2 and eC-A3. In the B branch, eC-A2 and eC-B3 show similar interactions, as these amino acids are conserved between PsaA and PsaB. **b**, Loops A-hi of PsaA (turquoise) and B-hi of PsaB (pink) surround the inter-polypeptide iron sulphur cluster F_x , coordinated by cysteines A578, A587 and B565, B574. Pseudo-C₂ axis as grey

vertical line. **c**, Phylloquinone Q_k -A (yellow) in its binding pocket. Q_k -B bound similarly by PsaB. Main chain of PsaA as ribbon and a carotenoid (car) in grey. Trp A697 π – π stacks with Q_k -A. Only one hydrogen bond, Leu A722 NH to carbonyl-oxygen 4 of Q_k -A, is formed. A carotenoid (car) is located in the vicinity of Q_k -A. **d**, View of the quinone binding site normal to the quinone plane.

In the three putative red-shifted dimers, the chlorin planes are nearly parallel, π -stacked at 3.5 Å interplanar distance and laterally shifted with Mg^{2+} – Mg^{2+} separation of 7.6–8.9 Å. This suggests π – π interactions in addition to the calculated strong excitonic coupling due to interaction of the Q_y transition dipoles.

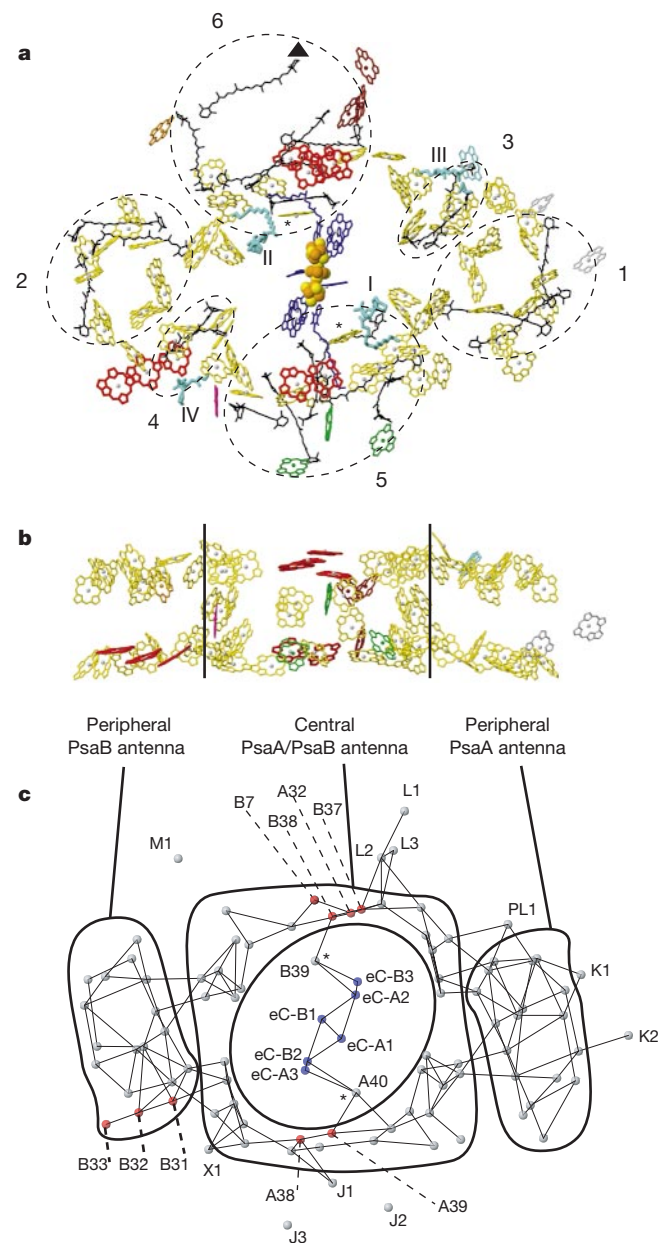


Figure 5 Spatial organization of the cofactors of the ETC and the antenna system in one monomer of PSI. **a**, View from the stromal side onto the membrane plane. The C₃ axis is indicated (triangle). Ring substituents of all chlorophylls omitted for clarity, putative nine excitonically coupled 'red' chlorophylls in red, chlorophylls coordinated by PsaA/PsaB in yellow, those bound to peripheral subunits in the colour of the coordinating subunits (Fig. 1), chlorophylls of the ETC in blue. Asterisks denote 'connecting chlorophylls' (see text). The 22 carotenoids are arranged in clusters 1 to 6. Four lipids (turquoise) labelled with roman numerals. **b**, Side view of PSI antenna rotated 90° about the horizontal axis with respect to **a**. Two vertical lines indicate separation of the antenna Chls coordinated by PsaA/B into a central part and two peripheral parts. **c**, Schematic view of the chlorophyll positions represented by their central Mg²⁺ ions; view direction as in **a**. All Mg²⁺–Mg²⁺ distances < 16 Å, thin lines. Chls belonging to the ETC in blue, those of the antenna in grey and red. Some of the chlorophylls referred to in the text are named by a capital letter indicating the subunits involved in axial liganding and numbered from N to C; PL1 is the single Chla bound to a phospholipid. The prefix aC- is omitted for clarity.

As the Chla pair aC-A32/aC-B7 (Fig. 5) is located at the luminal side of the 'trimerization domain' and forms hydrophobic contacts to the C-terminal region of PsaL of the adjacent monomer, it could have a role in excitation energy transfer between the monomers. These Chla molecules may be identical to chlorophylls absorbing at 719 nm, which are lost when the trimer of *S. elongatus* is split into monomers⁴⁰.

The two other Chla dimers aC-A38/aC-A39 and aC-B37/aC-B38 are coordinated to PsaA and PsaB at the stromal side of the membrane. They obey the pseudo-C₂ symmetry, with one dimer located close to PsaF and the other close to PsaL. Each of the two Chla pairs is close (centre-to-centre distance 16 Å) to one of the 'connecting' Chla molecules, aC-A38/aC-A39 to aC-A40 and aC-B37/aC-B38 to aC-B39. Theoretical and mutagenesis studies should be used to show whether these dimers play a role in fast excitation energy transfer to the ETC through the two connecting chlorophylls.

The carotenoids

The 22 carotenoids identified in the electron density were modelled as β -carotene, which is dominant in PSI^{43,44}. Of the carotenoids that were modelled as complete β -carotenes, 16 adopt the all-*trans* configuration and five contain one or two *cis* bonds, of which two were modelled as 9-*cis* and one each as 9,9'-*cis*, as 9,13'-*cis* and as 13-*cis* carotenoids, respectively. The numbers and configurations of these isomers agree well with results from high-performance liquid chromatography analysis on PSI from *S. vulcanus*⁴⁵. The carotenoids are deeply inserted in the membrane, with only few closer to the luminal or stromal surface. Their arrangement may be subdivided into six clusters (Fig. 5a). Clusters (1), (2), (3) and (4) contain 10 carotenoids and are bound to PsaA and PsaB by hydrophobic interactions, with (1) forming a few additional contacts to PsaK. Clusters (5) and (6), each consisting of six carotenoids, are in contact with PsaA and PsaB; (5) is also bound to PsaF and PsaJ and (6) to PsaI, PsaL and PsaM. The carotenoids in (1) and (3) are roughly related by the pseudo-C₂ axis to those in (2) and (4), respectively, but the pseudosymmetry is less obvious for (5) and (6).

The carotenoids might serve two major functions in photosynthesis (for review see ref. 46), namely light harvesting in the 450–570-nm range where absorption by Chla is negligible, and photoprotection by quenching excited Chla triplet states by a charge transfer mechanism, preventing formation of toxic singlet oxygen. The 22 carotenoids are in van der Waals contact (< 3.6 Å) to 60 Chla headgroups, thereby facilitating efficient energy transfer from excited carotenoids to the Chla molecules as well as quenching of Chla triplets. Several of the carotenoids show extended π – π stacking with chlorophylls, which appears ideal for fine-tuning the absorption properties of the antenna.

Lipids

Four lipids are found associated with PsaA/PsaB. Three are phospholipids (I, III, IV) and one is a galactolipid (II). These types of lipid were also identified biochemically (D. Linke, personal communication) and immunologically in PSI to be phosphatidylglycerol and monogalactosyldiglyceride^{44,47}. The head groups of the four lipids are well defined in the electron density map. They are located at the stromal side of the membrane plane and form hydrogen bonds with PsaA/B; the phosphodiester group of lipid (III) binds the antenna Chla aC-PL1. The two fatty acid chains of each lipid are anchored between transmembrane α -helices of PsaA/B and extend towards the middle of the membrane. Palmitic acid chains of 16 C atoms and stearic acid chains of 18 C atoms were modelled into the electron density map for the innermost lipids (I) and (II), respectively, whereas only shorter chains could be modelled for lipids (III) and (IV), probably owing to a higher degree of disorder. The lipids are related by the pseudo-C₂ symmetry: galactolipid (II)

and phospholipid (I) are close to the palisade of five transmembrane α -helices each in PsaB and PsaA (A/B-g to A/B-k), respectively. Of the other two lipids, one (III; bound to PsaA) is close to the monomer–monomer interface and the other (IV; bound to PsaB) is in contact with PsaX. The locations of lipids (I) and (II) close to the core of PSI and the binding of an antenna Chla by lipid (III) indicate that the four lipids are integral to and functionally important in the PSI complex and are not mere preparation artefacts.

Comparison of the antenna systems in PSI and PSII

The distribution of the peripheral antenna Chla molecules bound to PsaA and PsaB in PSI strongly resembles that of the membrane integral antenna proteins CP43 and CP47 of PSII. Their amino-acid sequences are similar to those of the N-terminal domains of PsaA and PsaB⁴⁸ and fold into comparable inner-membrane structures consisting of six transmembrane α -helices each³.

In both types of structures the Chla molecules are arranged in two quasi-two-dimensional layers. The shortest distances between Chlas of adjacent layers in PSI are >18 Å, whereas the inner-layer distances are less than 16 Å (Fig. 5c). As the centre-to-centre distance between Chla molecules determines, among other factors, the inter-molecule excitation transfer rates, it is more likely that an exciton is transferred within a single layer than that it is exchanged between Chlas in two different layers. It can be speculated that this might favour more efficient exciton transfer from these peripheral antennae to the reaction centre core rather than a comparable situation where the same number of cofactors would be evenly distributed over the whole depth of the membrane.

The PSII reaction centre domain formed by subunits D1 and D2 has a structure similar to the C-terminal domains of PsaA and PsaB and probably binds only two Chla molecules in addition to the cofactors of the ETC³. In contrast to PSII, the C-terminal domains of PsaA and PsaB of PSI cannot be considered as pure reaction centre domains⁴⁸, as they coordinate 12 (PsaA) and 13 (PsaB) Chlas of the antenna (Table 2). Moreover, the whole central part of the PSI core antenna, which contains 43 Chlas, is strikingly different from the corresponding region in PSII.

As each of the Chla molecules in the central PsaA/PsaB antenna of PSI is located at relatively short distances from the ETC, the excitation energy may be transferred with comparable efficiency to the cofactors of the ETC, in accordance with existing models⁴⁹. It will be of interest to clarify whether the almost complete lack of such a central antenna in PSII is associated with its lower efficiency, as compared to PSI, in transferring excitation energy to the reaction centre core⁵⁰.

This first structural model of PSI at atomic detail provides many new insights into the working of the photosynthetic system both of cyanobacteria and of plants. Whereas uncertainties remain as to the form adopted by and hence functioning of the peripheral PsaK, we have been able to resolve PsaX in the structure, proving it to be an intrinsic component of the photosystem. A number of novel modes of chlorophyll binding have also been revealed. The structure will help to answer current questions central to PSI chemistry, such as the role of the carotenoid cofactors and whether the PsaA/PsaB pseudosymmetry is reflected at the functional level. The structure provides a sound basis for functional analyses of exciton transfer and electron transport in plant and cyanobacterial photosystems, information that should prove valuable in a wealth of other contexts involving similar atomic interactions.

Received 26 February; accepted 2 May 2001.

- Nitschke, W. & Rutherford, A. W. Photosynthetic reaction centres: variations on a common structural theme? *Trends Biochem. Sci.* **16**, 241–245 (1991).
- Lancaster, C. R. D., Bibikova, M. V., Sabatino, P., Oesterhelt, D. & Michel, H. Structural basis of the drastically increased initial electron transfer rate in the reaction center from a *Rhodospseudomonas viridis* mutant described at 2.00-Å resolution. *J. Biol. Chem.* **275**, 39364–39368 (2000).
- Zouni, A. *et al.* Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution.

Nature **409**, 739–743 (2001).

- Krauss, N. *et al.* Photosystem I at 4 Å resolution represents the first structural model of a joint photosynthetic reaction centre and core antenna system. *Nature Struct. Biol.* **3**, 965–973 (1996).
- Klukas, O. *et al.* Localization of two phylloquinones, Q_K and Q_K', in an improved electron density map of photosystem I at 4 Å resolution. *J. Biol. Chem.* **274**, 7361–7367 (1999).
- Boekema, E. J. *et al.* Evidence for a trimeric organization of the photosystem I complex from the thermophilic cyanobacterium *Synechococcus* sp. *FEBS Lett.* **217**, 283–286 (1987).
- Brettel, K. Electron transfer and redox-cofactors in photosystem I. *Biochim. Biophys. Acta* **1318**, 322–373 (1997).
- Fromme, P. & Witt, H. T. Improved isolation and crystallization of photosystem I for structural analysis. *Biochim. Biophys. Acta* **1365**, 175–184 (1998).
- Mühlenhoff, U., Haehnel, W., Witt, H. T. & Herrmann, R. G. Genes encoding eleven subunits of photosystem I from the thermophilic cyanobacterium *Synechococcus* sp. *Gene* **127**, 71–78 (1993).
- Koike, K., Ikeuchi, M., Hiyaama, T. & Inoue, Y. Identification of photosystem I components from the cyanobacterium *Synechococcus vulcanus* by N-terminal sequencing. *FEBS Lett.* **253**, 257–263 (1989).
- Ikeuchi, M., Nyhus, K. J., Inoue, Y. & Pakrasi, H. B. Identities of four low-molecular-mass subunits of the photosystem I complex from *Anabaena variabilis* ATCC 29413. Evidence for the presence of the *psaI* gene product in a cyanobacterial complex. *FEBS Lett.* **287**, 5–9 (1991).
- Fromme, P., Schubert, W.-D. & Krauss, N. Structure of photosystem I: Suggestions on the docking sites for plastocyanin, ferredoxin and the coordination of P700. *Biochim. Biophys. Acta* **1187**, 99–105 (1994).
- Sun, J. *et al.* Oxidizing side of the cyanobacterial photosystem I. Evidence for interaction between the electron donor proteins and a luminal surface helix of the PsaB subunit. *J. Biol. Chem.* **274**, 19048–19054 (1999).
- Xu, Q., Yu, L., Chitnis, V. P. & Chitnis, P. R. Function and organization of photosystem I in a cyanobacterial mutant strain that lacks PsaF and PsaJ subunits. *J. Biol. Chem.* **269**, 3205–3211 (1994).
- Hippler, M., Drepper, F., Rochaix, J. D. & Mühlenhoff, U. Insertion of the N-terminal part of PsaF from *Chlamydomonas reinhardtii* into photosystem I from *Synechococcus elongatus* enables efficient binding of algal plastocyanin and cytochrome *c₆*. *J. Biol. Chem.* **274**, 4180–4188 (1999).
- Mühlenhoff, U. *et al.* Characterization of a redox-active cross-linked complex between cyanobacterial photosystem I and its physiological acceptor flavodoxin. *EMBO J.* **15**, 488–497 (1996).
- Adman, E. T., Sieker, L. C. & Jensen, L. H. The structure of a bacterial ferredoxin. *J. Biol. Chem.* **248**, 3987–3996 (1973).
- Falzone, C. J., Kao, Y. H., Zhao, J., Bryant, D. A. & Lecomte, J. T. Three-dimensional solution structure of PsaE from the cyanobacterium *Synechococcus* sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry* **33**, 6052–6062 (1994).
- Mayer, K. L., Shen, G., Bryant, D. A., Lecomte, J. T. & Falzone, C. J. The solution structure of photosystem I accessory protein E from the cyanobacterium *Nostoc* sp. strain PCC 8009. *Biochemistry* **38**, 13736–13746 (1999).
- Deisenhofer, J., Epp, O., Sinning, I. & Michel, H. Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodospseudomonas viridis*. *J. Mol. Biol.* **246**, 429–457 (1995).
- Scheer, H. *Chlorophylls* (CRC Press, Boca Raton, Florida, 1991).
- Watanabe, T., Kobayashi, M., Hongu, A., Nakazato, M. & Hiyaama, T. Evidence that a chlorophyll *a*' dimer constitutes the photochemical reaction centre I (P700) in photosynthetic apparatus. *FEBS Lett.* **235**, 252–256 (1985).
- Artz, K. *et al.* Relationship between the oxidation potential and electron spin density of the primary electron donor in reaction centers from *Rhodobacter sphaeroides*. *Proc. Natl Acad. Sci. USA* **94**, 13582–13587 (1997).
- Webber, A. N. *et al.* Site-directed mutations affecting the spectroscopic characteristics and midpoint potential of the primary donor in photosystem I. *Biochemistry* **35**, 12857–12863 (1996).
- Käss, H., Fromme, P., Witt, H. T. & Lubitz, W. Orientation and electronic structure of the primary donor radical cation P700^{•+} in Photosystem I: a single crystals EPR and ENDOR study. *J. Phys. Chem. B* **105**, 1225–1239 (2001).
- Kamrowski, A. *et al.* The quinone acceptor A₁ in photosystem I: Binding site, and comparison to Q_A in purple bacteria reaction centers. *J. Phys. Chem. B* **102**, 8278–8287 (1998).
- Iwaki, M. & Itoh, S. Structure of the phylloquinone-binding (Q_A) site in green plant photosystem I reaction centers: the affinity of quinones and quinoid compounds for the Q_A site. *Biochemistry* **30**, 5347–5352 (1991).
- Golbeck, J. H. In *Advances in Photosynthesis: The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 319–360 (Kluwer Academic, Dordrecht, Netherlands, 1994).
- Bittl, R., Zech, S. G., Fromme, P., Witt, H. T. & Lubitz, W. Pulsed EPR structure analysis of photosystem I single crystals: localization of the phylloquinone acceptor. *Biochemistry* **36**, 12001–12004 (1997).
- Yang, F. *et al.* Deletion of the PsaF polypeptide modifies the environment of the redox active phylloquinone (A₁). Evidence for unidirectionality of electron transfer in Photosystem I. *J. Phys. Chem. B* **102**, 8288–8299 (1998).
- Joliet, P. & Joliet, A. In vivo analysis of the electron transfer within photosystem I: are the two phylloquinones involved? *Biochemistry* **38**, 11130–11136 (1999).
- Guergova-Kuras, M., Boudreaux, B., Joliet, A., Joliet, P. & Redding, K. Evidence for two active branches for electron transfer in Photosystem I. *Proc. Natl Acad. Sci. USA* **98**, 4437–4442 (2001).
- Moser, C. C. & Dutton, P. L. Engineering protein structure for electron transfer function in photosynthetic reaction centers. *Biochim. Biophys. Acta* **1101**, 171–176 (1992).
- Schlodder, E., Falkenberg, K., Gergeleit, M. & Brettel, K. Temperature dependence of forward and reverse electron transfer from A₁–, the reduced secondary electron acceptor in photosystem I. *Biochemistry* **37**, 9466–9476 (1998).
- Fish, L., Kück, U. & Bogorad, L. Two partially homologous adjacent light-inducible maize chloroplast genes encoding polypeptides of the P700 chlorophyll *a* protein complex of photosystem I. *J. Biol. Chem.* **260**, 1413–1421 (1985).
- Chamrowsky, S. K. & Cammack, R. Direct determination of the midpoint potential of the acceptor X in chloroplast photosystem I by electrochemical reduction and ESR spectroscopy. *Photobiophys.* **4**, 195–200 (1982).
- Zhao, J., Li, N., Warren, P. V., Golbeck, J. H. & Bryant, D. A. Site-directed conversion of a cysteine to aspartate leads to the assembly of a [3Fe-4S] cluster in PsaC of photosystem I. The photoreduction of F_A is independent of F_B. *Biochemistry* **31**, 5093–5099 (1992).

38. Golbeck, J. H. A comparative analysis of the spin state distribution of in vitro and in vivo mutants of PsaC. *Photosynth. Res.* **61**, 107–144 (1999).
39. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* **2**, 55–75 (1948).
40. Pålsson, L. O. *et al.* Energy transfer and charge separation in photosystem I: P700 oxidation upon selective excitation of the long-wavelength antenna chlorophylls of *Synechococcus elongatus*. *Biophys. J.* **74**, 2611–2622 (1998).
41. Fleming, G. R. & van Grondelle, R. Femtosecond spectroscopy of photosynthetic light-harvesting systems. *Curr. Opin. Struct. Biol.* **7**, 738–748 (1997).
42. Kratky, C. & Dunitz, J. D. Ordered aggregation states of chlorophyll and some of its derivatives. *J. Mol. Biol.* **113**, 431–442 (1977).
43. Coufal, J., Hladik, J. & Sofrova, D. The carotenoid content of photosystem I-pigment-protein-complexes of the Cyanobacterium *Synechococcus elongatus*. *Photosynthetica* **23**, 603–616 (1989).
44. Makewicz, A., Radunz, A. & Schmidt, G. H. Comparative immunological detection of lipids and carotenoids on peptides of photosystem I from higher plants and cyanobacteria. *Z. Naturforsch.* **51c**, 319–328 (1996).
45. Ashikawa, I., Miyata, A., Koike, H., Inoue, Y. & Koyama, Y. Light-induced structural change of β -carotene in thylakoid membranes. *Biochemistry* **25**, 6154–6160 (1986).
46. Cogdell, R. J. Carotenoids in photosynthesis. *Pure Applied Chem.* **57**, 723–728 (1985).
47. Makewicz, A., Radunz, A. & Schmidt, G. H. in *Plant Lipid Metabolism* (eds Kader, J. -C. & Mazliak, P.) 156–160 (Kluwer Academic, Dordrecht, Netherlands, 1995).
48. Schubert, W. D. *et al.* A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. *J. Mol. Biol.* **280**, 297–314 (1998).
49. Valkunas, L., Liuolia, V., Dekker, J. P. & van Grondelle, R. Description of energy migration and trapping in photosystem I by a model with two distance scaling parameters. *Photosyn. Res.* **43**, 149–154 (1995).
50. Karapetyan, N. V., Holzwarth, A. R. & Rögner, M. The photosystem I trimer of cyanobacteria: molecular organization, excitation dynamics and physiological significance. *FEBS Lett.* **460**, 395–400 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

These studies are financially supported by Deutsche Forschungsgemeinschaft, by Bundesministerium für Bildung und Forschung and by Fonds der Chemischen Industrie. We thank D. DiFiore and H. Schmidt for technical assistance; W. Schroeder for supporting the analytical characterisation of the PSI preparations; B. Rasmussen and J. Lescar for long-term support during data collection at beamline ID2B at ESRF in Grenoble; and J. Hughes for reading and improving the manuscript. We particularly thank D. A. Bryant for discussions regarding PsaX and acknowledge R. Bittl, K. Brettel, M. Byrdin, P. Chitnis, J. Golbeck, B. Loll, W. Lubitz, E. Schlodder and D. Stehlik for discussions.

Correspondence and requests for materials should be addressed to P.F. (e-mail: fromme@phosis1.chem.tu-berlin.de) or N.K. (e-mail: nkrauss@chemie.fu-berlin.de). The X-ray crystallographic coordinates will be deposited at Brookhaven data bank under accession number 1JB0.

Skyrmions in a ferromagnetic Bose–Einstein condensate

Usama Al Khawaja & Henk Stoof

Institute for Theoretical Physics, University of Utrecht, Princetonplein 5,
3584 CC Utrecht, The Netherlands

Multi-component Bose–Einstein condensates^{1–3} provide opportunities to explore experimentally the wealth of physics associated with the spin degrees of freedom^{4–7}. The ground-state properties^{8–11} and line-like vortex excitations^{8,12,13} of these quantum systems have been studied theoretically. In principle, nontrivial spin textures consisting of point-like topological excitations, or skyrmions^{14,15}, could exist in a multi-component Bose–Einstein condensate, owing to the superfluid nature of the gas. Although skyrmion excitations are already known in the context of nuclear physics and the quantum-Hall effect, creating these excitations in an atomic condensate would offer an opportunity to study their physical behaviour in much greater detail, while also enabling an *ab initio* comparison between theory and experiment. Here we investigate theoretically the stability of skyrmions in a fictitious spin-1/2 condensate of ⁸⁷Rb atoms. We find that skyrmions can exist in such a gas only as a metastable state, but with a lifetime comparable to (or even longer than) the typical lifetime of the condensate itself.

An essential feature of a spinor Bose–Einstein condensate is that two or more hyperfine states of the atoms in the condensate have almost the same energy. As a result, this spin degree of freedom becomes a relevant dynamical variable, which gives rise to new excitations that are not present in the usual single-component Bose–Einstein condensates, where the spins are effectively frozen. One of these excitations is the skyrmion, which is a topological nontrivial spin texture. Roughly speaking, the skyrmion is a point-like object that can be created out of the ground state, in which all the spins are aligned, by reversing the average spin in a finite region of space. Although topological considerations indeed allow for these excitations, which we note are fundamentally different from the topologically trivial coreless vortices discussed in ref. 8, we need to know whether such a configuration is also energetically stable. However, when solving the appropriate Gross–Pitaevskii equation⁸, we find that, in equilibrium, the energy is always minimized by collapsing the skyrmion to zero size. Fortunately, it turns out that for sufficiently small sizes of the skyrmion a nonequilibrium stability mechanism starts to work. A number of atoms in the centre of the skyrmion, which we denote here as the core atoms, will be trapped by an effective three-dimensional potential barrier, that is, a repulsive shell with a finite radius that is induced by the gradients in the spin texture of the skyrmion itself. As the skyrmion shrinks in size, the barrier height of the repulsive shell increases and the radius decreases. This leads to a squeezing of the core atoms and thus to an increase in their energy, which ultimately stabilizes the skyrmion. As mentioned, this is not an equilibrium state of the condensate because the core atoms will tunnel over the barrier and give the skyrmion a finite lifetime. We note that our calculations are for a uniform condensate. However, our results are also valid for trapped gases, as the size of the skyrmion always turns out to be of the order of the correlation length of the condensate, which is typically much smaller than the size of the condensate. Moreover, in order to produce realistic estimates that can be compared with future experiments, we discuss only the case of a spin-1/2 ⁸⁷Rb condensate, because the spin-1 ²³Na condensate that has also been realized experimentally has an antiferromagnetic ground state.

We thus consider a uniform condensate of constant density n and

constant spinor $\zeta_z = (1, 0)$, with all spins being oriented in the positive z -direction. This represents the ground state of the gas. For the skyrmion, however, both n and ζ are position-dependent. A convenient way to take the position dependence of $\zeta(\mathbf{r})$ into account is to express the spinor as $\zeta(\mathbf{r}) = \exp\{-i2\mathbf{\Omega}(\mathbf{r})\cdot\hat{\mathbf{S}}\}\zeta_z$, where $\hat{\mathbf{S}}$ are the usual angular momentum operators for spin-1/2. The physical meaning of this formula is that the average spin at a position $\mathbf{r}$ is rotated by an angle $2\mathbf{\Omega}(\mathbf{r})$ from its initial orientation with $\mathbf{\Omega}(\mathbf{r})/\Omega(\mathbf{r})$ the axis of rotation. An explicit form of $\mathbf{\Omega}(\mathbf{r})$ now determines a specific texture of the skyrmion. We consider only the most symmetric shape of the skyrmion, because on general grounds this is expected to have the lowest energy. We thus take $\mathbf{\Omega}(\mathbf{r}) = \omega(r)\mathbf{r}/r$. The boundary conditions at $r = 0$ and $r \rightarrow \infty$ that the function $\omega(r)$ should satisfy are the following. First, at $r \rightarrow \infty$ all spins must be oriented as in the ground state, since otherwise it requires an infinite amount of energy to create the skyrmion. This implies $\lim_{r \rightarrow \infty} \omega(r) = 0$. Along the z -axis the spins are also not rotated by our ansatz for $\mathbf{\Omega}(\mathbf{r})$, so in order to have a nonsingular texture of the spinor with a non-zero winding number, we must require that $\omega(0) = 2\pi$. Finally, to avoid a singular behaviour of $\mathbf{\Omega}(\mathbf{r})$ itself, we take only functions $\omega(r)$ with zero slope at the origin. In summary, $\omega(r)$ is therefore a monotonically decreasing function that starts from 2π at the origin and reaches zero when $r \rightarrow \infty$. The specific functional form of $\omega(r)$ at intermediate distances between zero and infinity is not crucial for the stability of the skyrmion. Only the above boundary conditions are important for that.

From a quantum mechanical point of view, the condensate is described by a macroscopic wavefunction, or order parameter, $\psi(\mathbf{r}) \equiv \sqrt{n(\mathbf{r})}\zeta(\mathbf{r})$. Furthermore, the grand-canonical energy (energy – number of atoms times chemical potential) of the gas can in the usual mean-field approximation be expressed as a function of $n(\mathbf{r})$ and $\zeta(\mathbf{r})$ ⁸. In detail, we have:

$$E[n(\mathbf{r}), \zeta(\mathbf{r})] \equiv \int d\mathbf{r} \left[\frac{\hbar^2}{2m} (\nabla \sqrt{n(\mathbf{r})})^2 - \mu n(\mathbf{r}) + \frac{\hbar^2}{2m} n(\mathbf{r}) |\nabla \zeta(\mathbf{r})|^2 + \frac{1}{2} T^{2B} n^2(\mathbf{r}) \right] \quad (1)$$

where m is the mass of the atoms, μ is their chemical potential and $T^{2B} = 4\pi a \hbar^2/m$ is the appropriate coupling constant that represents the strength of the interatomic interactions in terms of the positive scattering length a . It is clear from the above expression that the gradients in the spin texture lead to a contribution to the energy density, which is proportional to $|\nabla \zeta(\mathbf{r})|^2$. Using the above-mentioned form of $\zeta(\mathbf{r})$ and inserting the explicit forms of the spin-1/2 matrices, the square of the spinor gradient can be written explicitly as:

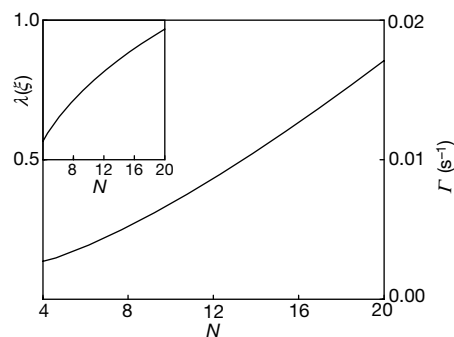


Figure 1 Decay rate Γ and size λ (inset) of the skyrmion as a function of the number of atoms trapped in the core. The two curves are calculated for the parameters of a ⁸⁷Rb spinor condensate with a density of 10^{13} cm^{-3} . The size is calculated in units of the correlation length ξ , which for these parameters equals $0.85 \mu\text{m}$.

$$|\nabla\zeta(\mathbf{r})|^2 = 2\left(\frac{\sin(\omega(r))}{r}\right)^2 + \left(\frac{d\omega(r)}{dr}\right)^2 \quad (2)$$

In principle, both $n(r)$ and $\omega(r)$ can now be calculated exactly by minimizing the energy functional in equation (1) with respect to arbitrary functions $n(r)$ and $\omega(r)$ that satisfy the boundary conditions. However, the resulting nonlinear and coupled equations are quite difficult to solve. Therefore, we use here a variational approach and take for $\omega(r)$ the simple ansatz $\omega(r) = 4 \cot^{-1}[(r/\lambda)^2]$, where the variational parameter λ physically corresponds to the size of the skyrmion. This ansatz is chosen here because it automatically incorporates the correct boundary conditions and leads to a minimal skyrmion energy as compared to other ansatzes that we have considered. Having specified $\omega(r)$, we then calculate $n(r)$ exactly by solving numerically the differential equation for $n(r)$ obtained by varying $E[n(\mathbf{r}), \zeta(\mathbf{r})]$ with respect to $n(r)$. Substituting this density profile back into the energy functional, the energy of the skyrmion becomes a function of λ only and the equilibrium properties can be obtained by minimizing this energy with respect to λ .

Inserting our ansatz for $\omega(r)$ in equation (2), $\hbar^2|\nabla\zeta(\mathbf{r})|^2/2m$ takes the shape of an off-centred potential barrier with a maximum of $24.3\hbar^2/2m\lambda^2$ located at 0.68λ . We now distinguish between two cases. The first case occurs when the height of the barrier is lower than the chemical potential μ of the atoms. In this case, atoms can move freely across the barrier and equilibrium is quickly achieved. In the second case, the barrier height is higher than the chemical potential and thus atoms become trapped behind the barrier near the centre of the skyrmion. Equating the barrier height with $\mu = T^{2B}n$ gives therefore the maximum value of $\lambda_{\max} \approx 5\xi$ below which trapping takes place. Here ξ is the correlation length given by $\xi = 1/\sqrt{8\pi an}$, and n is the density at $r \gg \lambda$. For $\lambda < \lambda_{\max}$ we calculate first the metastable size of the

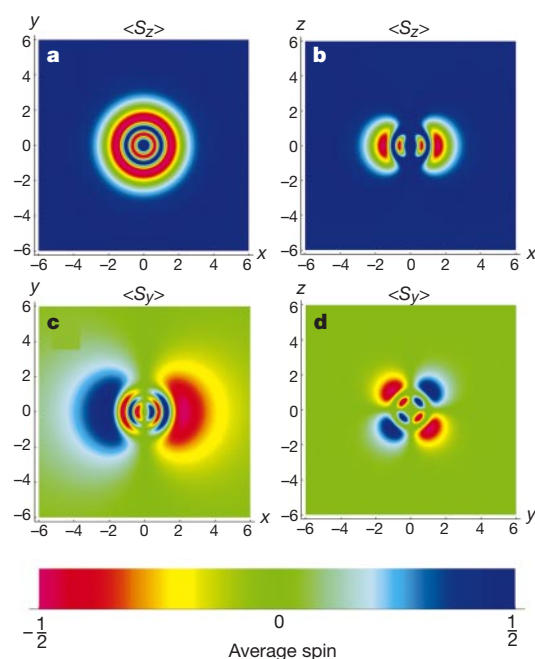


Figure 2 The average spin texture of the skyrmion. Shown are the components $\langle S_x(\mathbf{r}) \rangle = \zeta^+(\mathbf{r})\hat{S}_x\zeta(\mathbf{r})$ and $\langle S_y(\mathbf{r}) \rangle$ in the three cartesian planes. A value of $\lambda = \xi$, which corresponds to 20 core atoms, was chosen to compute these quantities. The distances are in units of the coherence length ξ . The $\langle S_x(\mathbf{r}) \rangle$ component can be obtained from the $\langle S_y(\mathbf{r}) \rangle$ component, by using the cylindrical symmetry of the skyrmion. **a, b**, It is clear that our skyrmion is composed of two coaxial tori. The two pink rings represent the cores of these two tori where $\langle S_z \rangle$ reaches its maximum negative value of $-1/2$. **c, d**, Show the cross-sections of the two tori.

skyrmion by minimizing for a fixed number of core atoms the total energy of the condensate. Next we calculate the tunnelling rate for the core atoms to escape to the outer region (see Methods). The results of these calculations are presented in Fig. 1, where we plot the metastable size of the skyrmion as a function of the number of the core atoms trapped by the texture barrier, together with the corresponding tunnelling rate. As mentioned previously, we use the parameters of ^{87}Rb with a density of $n = 10^{13} \text{ cm}^{-3}$. We see from Fig. 1 that only a few atoms in the core of the skyrmion are needed to stabilize it and to give it a sufficiently long lifetime. We note that the tunnelling rate is considerably less than the decay rate of the condensate, which is due to two-body collisions¹⁶ and equal to Gn with a measured value of $G \approx 2.2 \times 10^{-14} \text{ cm}^{-3} \text{ s}^{-1}$. Also, the lifetime becomes even larger for slightly smaller values of n .

Next, we address some of the interesting properties of the skyrmion, which are its texture and its dynamics. The texture can best be presented in terms of the three average spin components $\langle S_x(\mathbf{r}) \rangle \equiv \zeta^+(\mathbf{r})\hat{S}_x\zeta(\mathbf{r})$, $\langle S_y(\mathbf{r}) \rangle$, and $\langle S_z(\mathbf{r}) \rangle$ in the three cartesian planes. These are shown in Fig. 2. The most interesting dynamical property of the skyrmion comes from the fact that if all spins of the texture are rotated around the z axis by a constant angle θ , that is, $\zeta(\mathbf{r}) \rightarrow \exp(-i\theta\hat{S}_z)\zeta(\mathbf{r})$, the resulting skyrmion will have the same energy. As a result, the angle θ undergoes phase diffusion. It also leads to a Josephson-like coupling between two skyrmions, which will have important consequences for the physics of the skyrmion lattice¹⁷ that, as in an Abrikosov vortex lattice, will form at sufficiently low temperatures. Another dynamical property is the centre-of-mass motion of the skyrmion, which can be shown to be identical to that of a massive particle. As a result the skyrmion will be accelerated towards the edge of the condensate in a trapped situation.

Although skyrmion–antiskyrmion pairs can be created by the Kibble mechanism in a temperature quench or by sufficiently shaking up the condensate, a more controlled way of creating a skyrmion can be achieved by using a magnetic field configuration in which the fictitious magnetic field is always pointing radially outward and its magnitude increases monotonically from zero at the origin to a maximum value for large distances from the origin. Applying this field configuration for such a long time that the spins at large distances have precessed exactly twice around the local magnetic field, creates a single skyrmion. Of course, for a real magnetic field the above configuration requires the use of magnetic monopoles, but for a fictitious magnetic field it can be achieved by appropriately tailoring the detuning, the polarization, and the intensity of two pulsed Raman lasers. The required spatial dependence of the detuning can be created experimentally by separating

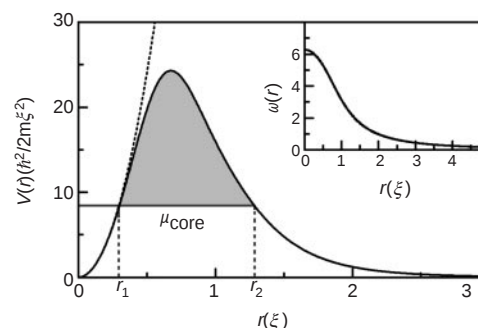


Figure 3 The potential barrier produced by the skyrmion texture. The value of λ used is again approximately ξ which corresponds to 20 core atoms. The lifetime of the skyrmion is determined by the tunnelling of core atoms with a chemical potential μ_{core} to the right of the barrier. The square root of the shaded area is the integrand of equation (4). The dotted curve represents the harmonic approximation to $V(r)$ for small r/λ . The inset shows $\omega(r)$ for the same value of λ .

the centres of the magnetic traps for the two spin species along the z axis⁶. Furthermore, the desired behaviour of the Rabi frequency can be achieved by making, with the first Raman laser, two standing waves in the x and y directions that are both polarized perpendicular to the z axis. For the other Raman laser we only require that it produces a travelling wave with a polarization that has a non-zero projection on the z axis, because we want to realize a $\Delta m = 0$ transition in this case. In the above geometry the skyrmion is created exactly in the plane where the detuning vanishes and in the nodes of the first Raman laser. We note that as the distance between these nodes is generally much bigger than the correlation length, we create in this manner a large skyrmion that will start to shrink but ultimately self-stabilizes at a smaller size. Once created, the skyrmion can be easily observed by the usual expansion experiments that have recently also been used to observe vortices¹⁸. As with vortex rings, we then observe an almost complete depletion of the condensate in a ring around the position of the skyrmion. □

Methods

To calculate the energy of the skyrmion we solve the equation for the density profile that is obtained from minimizing the energy functional in equation (1). We solve this equation numerically for the region outside the core. Inside the core we solve the equation analytically by using the Thomas–Fermi approximation, which amounts to neglecting the gradients of the density profile. Using our ansatz for $\omega(r)$, the texture gradient potential $V(r) = \hbar^2 |\nabla \xi(r)|^2 / 2m$ reads:

$$V(r) = \frac{\hbar^2}{2m \lambda^2} \frac{32 (r/\lambda)^2 [3 + 2(r/\lambda)^4 + 3(r/\lambda)^8]}{[1 + (r/\lambda)^4]^4} \quad (3)$$

For small r/λ this potential can be approximated by a harmonic potential with a characteristic frequency $\omega_0 = \sqrt{96 \hbar^2 / 2m \lambda^2} \lambda^{-1}$ and width $l = \lambda / \sqrt{96}$, as shown in the dotted curve in Fig. 3. Specifically, the use of a Thomas–Fermi approximation is justified when the ratio $2Na/l = 2\sqrt{96}Na/\lambda$ is bigger than 1. From Fig. 1, we observe that this ratio equals approximately 1 for $N = 4$ and increases for larger N .

The lifetime of the skyrmion is estimated by calculating the tunnelling rate from the core to the outer region over the barrier $V(r)$. To this end we employ the following WKB (Wentzel, Kramers and Brillouin) expression for the tunnelling rate¹⁹:

$$\Gamma = \frac{\omega_0}{2\pi} \exp \left[-2 \int_{r_1}^{r_2} dr \sqrt{\frac{2m}{\hbar^2} (V(r) - \mu_{\text{core}})} \right] \quad (4)$$

where μ_{core} is the chemical potential of the core atoms. The radial points r_1 and r_2 are the points where $V(r)$ and μ_{core} intersect, as shown in Fig. 3. The chemical potential μ_{core} is calculated by differentiating the total energy of the core with respect to the number of core atoms.

Received 15 November 2000; accepted 20 April 2001.

- Myatt, C. J., Burt, E. A., Ghrist, R. W., Cornell, E. A. & Wieman, C. E. Production of two overlapping Bose–Einstein condensates by sympathetic cooling. *Phys. Rev. Lett.* **78**, 586–589 (1997).
- Stamper-Kurn, D. M. *et al.* Optical confinement of a Bose–Einstein condensate. *Phys. Rev. Lett.* **80**, 2027–2030 (1998).
- Stenger, J. *et al.* Spin domains in ground-state Bose–Einstein condensates. *Nature* **396**, 345–348 (1998).
- Hall, D. S., Matthews, M. R., Ensher, J. R., Wieman, C. E. & Cornell, E. A. Dynamics of component separation in a binary mixture of Bose–Einstein condensates. *Phys. Rev. Lett.* **81**, 1539–1542 (1998).
- Stamper-Kurn, D. M. *et al.* Quantum tunneling across spin domains in a Bose condensate. *Phys. Rev. Lett.* **83**, 661–664 (1999).
- Matthews, M. R. *et al.* Watching a superfluid untwist itself: Recurrence of Rabi oscillations in a Bose–Einstein condensate. *Phys. Rev. Lett.* **83**, 3358–3361 (1999).
- Matthews, M. R. *et al.* Vortices in a Bose–Einstein condensate. *Phys. Rev. Lett.* **83**, 2498–2501 (1999).
- Ho, T.-L. Spinor Bose condensates in optical traps. *Phys. Rev. Lett.* **81**, 742–745 (1998).
- Ohmi, T. & Machida, K. Bose–Einstein condensation with internal degrees of freedom in alkali atom gases. *J. Phys. Soc. Jpn* **67**, 1822–1825 (1998).
- Law, C. K., Pu, H. & Bigelow, N. B. Quantum spins mixing in spin Bose–Einstein condensates. *Phys. Rev. Lett.* **81**, 5257–5261 (1998).
- Ho, T.-L. & Yip, S.-K. Fragmented and single condensate ground states of spin-1 Bose gas. *Phys. Rev. Lett.* **84**, 4031–4034 (2000).
- Yip, S.-K. Internal vortex structure of a trapped spinor Bose–Einstein condensate. *Phys. Rev. Lett.* **83**, 4677–4681 (1999).
- Williams, J. E. & Holland, M. J. Preparing topological states of a Bose–Einstein condensate. *Nature* **401**, 568–572 (1999).
- Skyrme, T. H. R. A non-linear field theory. *Proc. R. Soc. Lond. A* **260**, 127–138 (1961).
- Skyrme, T. H. R. A unified field theory of mesons and baryons. *Nucl. Phys.* **31**, 556–569 (1962).
- Julienne, P. S., Mies, F. H., Tiesinga, E. & Williams, C. J. Collisional stability of double Bose condensates. *Phys. Rev. Lett.* **78**, 1880–1883 (1997).
- Côté, R. *et al.* Collective excitations, NMR, and phase transitions in Skyrme crystals. *Phys. Rev. Lett.* **78**, 4825–4828 (1997).

- Madison, K. W., Chevy, F., Wohlleben, W. & Dalibard, J. Vortex formation in a stirred Bose–Einstein condensate. *Phys. Rev. Lett.* **84**, 806–809 (2000).
- Stoof, H. T. C. Macroscopic quantum tunneling of a Bose–Einstein condensate. *J. Stat. Phys.* **87**, 1353–1366 (1997).

Acknowledgements

We thank M. Bijlsma for help in the numerical calculations and for helpful remarks. We also thank J. Anglin, G. 't Hooft, D. Olive and J. Smit for useful discussions. This work is supported by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Correspondence and requests for materials should be addressed to U.A.K. (e-mail: u.alkhawaja@phys.uu.nl).

Interplay of magnetism and high- T_c superconductivity at individual Ni impurity atoms in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$

E. W. Hudson^{†*}, K. M. Lang^{*}, V. Madhavan^{*}, S. H. Pan^{‡,§}, H. Eisaki^{§||}, S. Uchida[§] & J. C. Davis^{*}

^{*} Department of Physics, University of California, Berkeley, California 94720, USA

[†] National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

[‡] Department of Physics, Boston University, Boston, Massachusetts 02215, USA

[§] Department of Superconductivity, University of Tokyo, Yayoi, 2-11-16 Bunkyo-ku, Tokyo 113-8656, Japan

^{||} Department of Applied Physics, Stanford University, Stanford, California, 94205-4060, USA

Magnetic interactions and magnetic impurities are destructive to superconductivity in conventional superconductors¹. By contrast, in some unconventional macroscopic quantum systems (such as superfluid ³He and superconducting UGe₂), the superconductivity (or superfluidity) is actually mediated by magnetic interactions. A magnetic mechanism has also been proposed for high-temperature superconductivity^{2–6}. Within this context, the fact that magnetic Ni impurity atoms have a weaker effect on superconductivity than non-magnetic Zn atoms in the high- T_c superconductors has been put forward as evidence supporting a magnetic mechanism^{5,6}. Here we use scanning tunnelling microscopy to determine directly the influence of individual Ni atoms on the local electronic structure of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. At each Ni site we observe two d-wave impurity states^{7,8} of apparently opposite spin polarization, whose existence indicates that Ni retains a magnetic moment in the superconducting state. However, analysis of the impurity-state energies shows that quasi-particle scattering at Ni is predominantly non-magnetic. Furthermore, we show that the superconducting energy gap and correlations are unimpaired at Ni. This is in strong contrast to the effects of non-magnetic Zn impurities, which locally destroy superconductivity⁹. These results are consistent with predictions for impurity atom phenomena^{5,6} derived from a magnetic mechanism.

In our studies we use two different $\text{Bi}_2\text{Sr}_2\text{Ca}(\text{Cu}_{1-x}\text{Ni}_x)_2\text{O}_{8+\delta}$ (BSCCO) single crystals, grown by the floating-zone technique. These crystals have $x = 0.005$ with $T_c = 83$ K and $x = 0.002$ with $T_c = 85$ K, respectively. The Ni atoms substitute for Cu atoms in the superconducting CuO_2 plane and are believed to be in the $\text{Ni}^{2+}3d^8$ electronic state, as compared to the $\text{Cu}^{2+}3d^9$ of Cu. Above the superconducting transition temperature T_c each Ni atom possesses a strong magnetic moment of around $1.5\mu_B$ (ref. 10). The samples

are cleaved in cryogenic ultra-high vacuum at 4.2 K to expose the BiO crystal plane and, when inserted into the scanning tunnelling microscope (STM) head, are imaged with atomic resolution.

For d-wave superconductors, theory predicts that quasi-particle scattering at an impurity atom creates a local electronic state (impurity state) nearby^{7,8,11–17}. Such states, which have been analysed for both potential (non-magnetic)^{7,8,11,12,14–16} and magnetic^{8,12–17} scattering, can be thought of as almost localized quantum orbitals of well defined energy Ω and spatial structure^{9,18}.

The spatial distribution of impurity states as Ni impurity atoms can be imaged by measuring, as a function of position, the differential tunnelling conductance $G = dI/dV$. This is proportional to the local density of states (LDOS) and we refer to this process as LDOS mapping. As an example, Fig. 1 shows two simultaneously acquired LDOS maps taken at sample bias $V = \pm 9$ mV. They reveal both the particle-like (positive bias) and hole-like (negative bias) components of one of the impurity

states that exist at each Ni. At +9 mV ‘+-shaped’ regions of higher LDOS are observed, whereas at -9 mV the corresponding higher LDOS regions are ‘X-shaped’. LDOS maps at $V = \pm 19$ mV show the particle-like and hole-like components of a second impurity state at Ni whose spatial structure is very similar to that at $V = \pm 9$ mV.

High-resolution ± 9 -mV LDOS maps (Fig. 2a, b) acquired simultaneously with a topographic image of the BiO surface (Fig. 2c) show in detail how an impurity state consists of two spatially complementary components. By this we mean that the particle-like LDOS is high in regions where the hole-like LDOS is low, and vice versa. To illustrate these relationships, Fig. 2d shows a schematic diagram of the relative locations of Cu atoms, the orientation of the $d_{x^2-y^2}$ superconducting order parameter (OP), and the location of the particle-like (green) and hole-like (purple) components of the impurity state.

We also probe the energy dependence of the LDOS at several

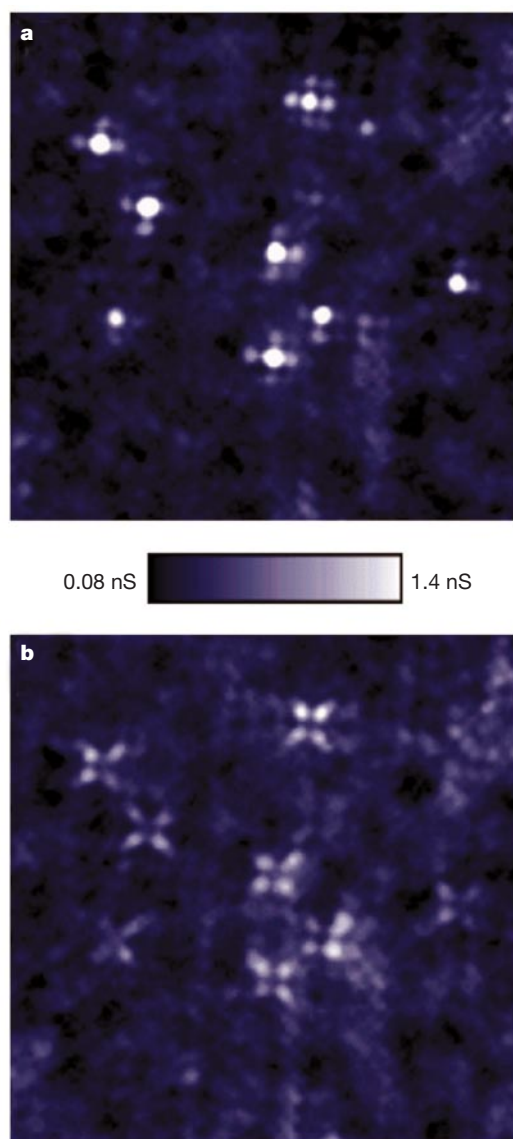


Figure 1 Maps of the local density of electronic states at energies $E = \pm 9$ meV, revealing the locations of the Ni impurity states. Two $128 \text{ \AA}$ square differential conductance maps of Ni-doped BSCCO. **a**, At sample bias +9 mV, showing the ‘+-shaped’ regions of high local density of states (LDOS) associated with the Ni atoms. **b**, At -9 mV, showing the 45° spatially rotated ‘X-shaped’ pattern. The differential conductance $G(V) = dI/dV$ measured at sample bias voltage V and tunnelling current I is proportional

to the LDOS at energy $E = eV$. In such an LDOS map, states associated with positive sample bias are referred to as particle-like whereas those at negative sample bias are hole-like. There are ~ 8 Ni atom sites in the field of view. As the areal densities of these phenomena in two crystals agree with the two different Ni dopant densities, we conclude that they are associated with Ni atoms. The junction resistance was $1 \text{ G}\Omega$ at -100 mV for the simultaneously obtained maps and all data were acquired at 4.2 K.

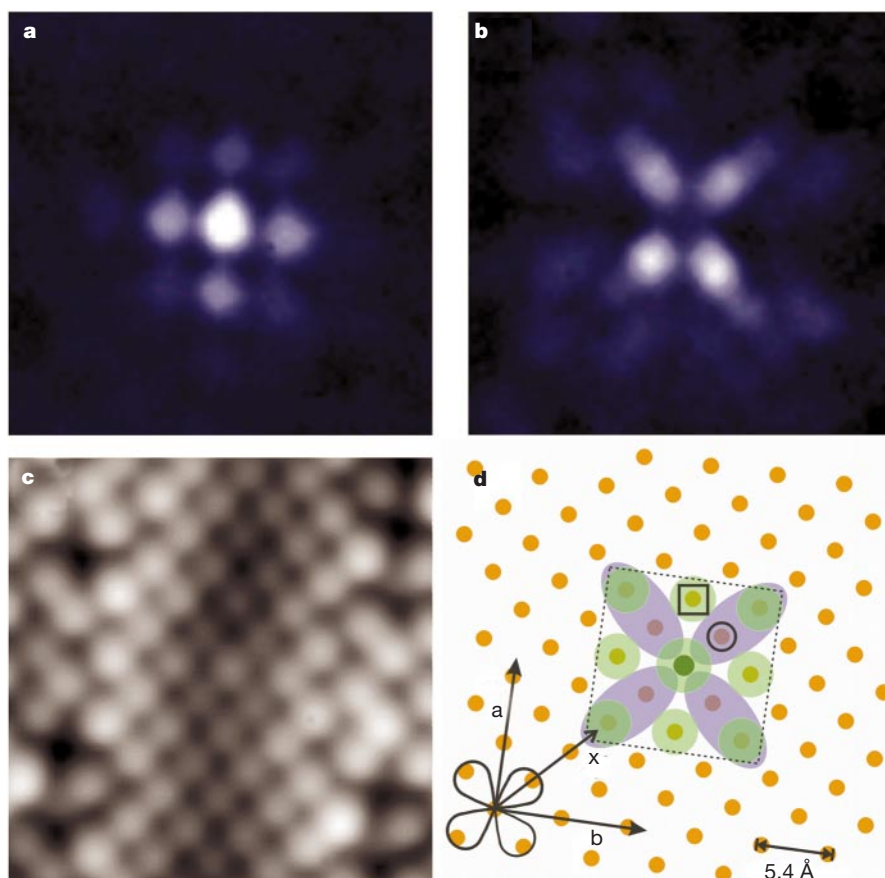


Figure 2 Detailed spatial structure of the impurity state at a single Ni atom. **a, b**, 35 Å square differential conductance maps above an Ni atom at **(a)** +9 mV and **(b)** -9 mV. **c**, A 35 Å square atomic resolution topograph of the BiO surface obtained simultaneously with the maps. **d**, Using the information in **a, b** and **c** and the fact that the Cu atoms in the CuO₂ plane lie directly beneath the Bi atoms in the BiO plane, we draw a schematic showing the relative position of the Ni atom (black solid circle) with respect to the Cu atoms (orange solid circles) in the invisible CuO₂ plane. Also depicted is the orientation of the $d_{x^2-y^2}$ order parameter with respect to the **a**- and **b**- axis, and the positions of the first NN (open

circle) and second NN (open square) atoms where the spectra shown in Fig. 3 are acquired. The ~16 Å square region inside which the spatial integration of the whole impurity-state LDOS was carried out is also shown. The junction resistance is 1 GΩ at -100 mV for all data. The bright central maximum of the '4-shaped' LDOS map always occurs at the site of a Bi surface atom, consistent with the Ni atom being at the Cu site ~5 Å below. The spatial structure in LDOS of an impurity state, when observed at the BiO plane, may not be identical to that in the CuO₂ plane¹⁶. However, the magnitude of the impurity-state energies should be independent of interlayer matrix element effects.

locations near a Ni atom. Figure 3a shows typical dI/dV spectra taken at four locations: above the Ni atom, above the first nearest neighbour (1-NN) Cu atom, above the second nearest neighbour (2-NN) Cu atom, and at a distance of 30 Å from the impurity site; and Fig. 3b shows the average spectrum for the whole impurity-state region. There are two clear particle-like LDOS peaks above the Ni site. The average magnitudes of these on-site impurity-state energies measured at eight different Ni sites are $\Omega_1 = 9.2 \pm 1.1$ meV and $\Omega_2 = 18.6 \pm 0.7$ meV. Both these peaks become hole-like at all the 1-NN Cu sites and again particle-like at the 2-NN Cu sites. Despite these complexities, the spatially averaged spectrum for this whole region (Fig. 3b) remains very close to particle-hole symmetric.

Several conclusions can be drawn from these data. First, the two on-site LDOS peaks reveal that there are two distinct impurity states associated with each Ni atom. This can be explained by theories that consider both potential and magnetic interactions⁸. In a d-wave superconductor, theory indicates that potential scattering generates a single spin-degenerate impurity state^{7,8,11}. A weak additional magnetic interaction between an impurity moment and the quasi-particle spin lifts the spin degeneracy, creating two spin-polarized impurity states at each magnetic impurity atom⁸. The good qualitative agreement between the data and this model indicates that Ni atoms do possess a magnetic moment in the superconducting state.

Second, as shown in Figs 2 and 3, the spectral weight of an impurity state oscillates between particle-like and hole-like as a function of distance from the Ni atom. Nevertheless, the average conductance spectrum over the whole impurity state remains almost particle-hole symmetric. Theoretically, such overall particle-hole symmetry in an impurity-state spectrum is expected when superconductivity is not disrupted^{12,14-16,19,20}. Furthermore, calculations for the d-wave impurity-state wavefunction show that the particle-like and hole-like LDOS should be spatially complementary^{8,12,14-16}. The agreement of our observations with these theoretical models shows that superconductivity is not disrupted locally by the magnetic moment of Ni.

A third observation is that the superconducting gap magnitude (as deduced from the energy of the coherence peaks) does not change as we approach the impurity site. This is shown in Fig. 4, which shows a series of conductance spectra as a function of distance from the Ni site. The particle-like coherence peak is depleted to provide spectral density for the impurity state, but the gap magnitude (28 mV) deduced from the hole-like coherence peak location is unperturbed.

In summary, two local states in excellent qualitative agreement with d-wave impurity scattering theory can be identified at each Ni impurity atom. The presence of two impurity states indicates that the Ni atom possesses a magnetic moment, whereas the existence

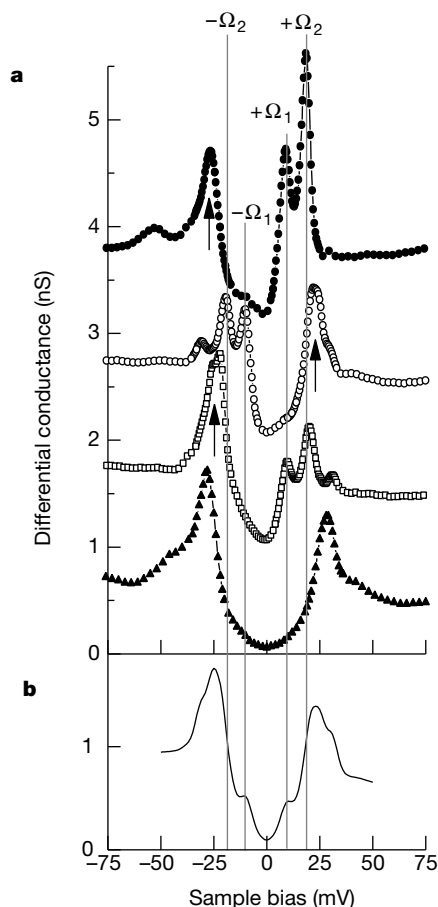


Figure 3 Differential conductance spectra above the Ni atom and at several nearby locations. Differential conductance spectra obtained at four positions near the Ni atom showing the maxima at $V = \pm \Omega_1/e$ and $V = \pm \Omega_2/e$ and the superconducting coherence peaks. These curves are shifted from each other by 1 nS for clarity. **a**, The spectrum above the Ni atom site (solid circles), above the first nearest neighbour Cu atom position (open circles), above the second nearest neighbour Cu atom (squares), and a typical spectrum 30 Å from an impurity atom (triangles). The location of the Ni atom and the first and second nearest neighbours is shown in Fig. 2. The spectra show rapid variations in LDOS magnitude as a function of position. They also reveal the persistence of the coherence peak (shown by the upward pointing arrows) in any given spectrum on the side opposite to the impurity-state resonance peaks. **b**, The average conductance spectrum for a ~ 16 Å square area centred on an Ni atom, showing the particle-hole symmetric nature of the spatially integrated LDOS. The junction resistance was 1 GΩ at -200 mV for all spectra.

and structure of the particle-like to hole-like oscillations and the constant gap magnitude indicate that superconductivity is nowhere disrupted.

We now turn to the implications of these data for proposals^{5,6} that differences between the influence of ‘magnetic’ Ni and ‘non-magnetic’ Zn impurity atoms are evidence of a magnetic mechanism for high-temperature superconductivity (HTSC). Previous experimental studies using Ni and Zn impurities have revealed two apparently contradictory aspects of their effects on HTSC. On one hand, resistivity^{21,22}, microwave surface resistivity²³ and T_c (refs 21, 22) measurements show similar dependences on Zn and Ni doping densities. On the other hand, probes sensitive to magnetic phenomena such as NMR^{24–30} and inelastic neutron scattering (INS)³¹, and probes of the superfluid density such as penetration depth²³ and muon-spin-rotation (μ SR)³², show marked differences between Ni- and Zn-doped samples.

Analysis of the Ni impurity-state energies using the model of ref. 8 is the first step towards understanding this situation. In that model, potential scattering is represented by an on-site energy U ,

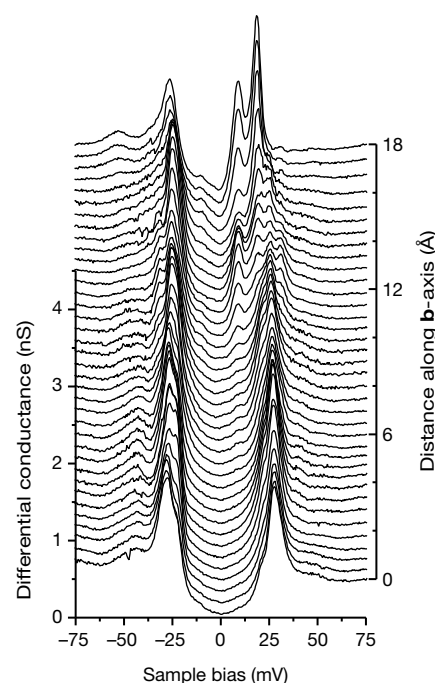


Figure 4 Evolution of the dI/dV spectra and superconducting energy gap as a function of distance from the Ni atom. A series of differential conductance spectra (shifted vertically from each other for clarity), taken as a function of distance away from the centre of a Ni impurity state along the **b**-axis. The top spectrum is taken above the Ni site. The spectra, each separated by 0.5 Å, show that over a ~ 20 Å range the magnitude of the superconducting gap $\Delta_0 = 28$ meV, as measured by the separation between the hole-like coherence peak and the Fermi level, is not suppressed in the presence of the Ni impurity. Line cuts approaching the Ni site from other directions lead to the same conclusion. There are variations in Δ_0 throughout the crystal but this study was carried out in a patch with constant Δ_0 of 28 meV. The junction resistance was 1 GΩ at -100 mV for all spectra.

and magnetic interaction is modelled as having energy $W = JS \cdot s$. Here S is the classical spin of the impurity atom, s is the quasi-particle spin and J is the exchange energy. The solution for the on-site resonance energy can be written approximately as⁸

$$\frac{\Omega_{1,2}}{\Delta_0} = \frac{-1}{2N_F(U \pm W) \ln[(8N_F(U \pm W))]}$$

Here Δ_0 is the maximum magnitude of the superconducting gap and N_F is the normal density of states per site at the Fermi energy. By substituting the measured values of Ω_1 and Ω_2 and $\Delta_0 = 28$ meV into this equation, we find $N_F U = -0.67$ and $|N_F W| = 0.14$. This represents a surprising new insight because Ni is usually regarded as a source of magnetic scattering, but here we identify the dominant effects as due to potential scattering.

The apparent conflicts between results from different probes may now be reconciled. The first set of probes is sensitive to scattering. Calculation of the potential scattering phase shift $\delta_0 = \tan^{-1}(\pi N_F U)$ for Ni gives $\delta_0 = 0.36\pi$, whereas our previous results showed that Zn is a unitary scatterer ($\delta_0 \approx \pi/2$)⁹. The similarity of these phase shifts imply that phenomena dependent on scattering should be quite similar in Ni- and Zn-doped samples. In fact, using these parameters in an Abrikosov–Gorkov model¹ (and ignoring Ni’s magnetic potential), we calculate that T_c would be suppressed only about 20% faster by Zn than by Ni, certainly within the range of experimental observations^{21,22}.

Among the second set of measurement techniques are those sensitive to superconductivity itself. For example, penetration depth²³ and μ SR³² measurements show that, in the bulk, Zn strongly depletes superfluid density but Ni has a much weaker impact. The STM data now provide a microscopic explanation for these differ-

ences—Zn atoms locally destroy superconductivity⁹ while Ni atoms do not.

Finally, magnetic probes sensitive to spin fluctuations reveal marked changes with Zn-doping^{24,26–31} but only weak perturbations with Ni-doping^{24,25,29,31}. Explanations for these phenomena have been proposed^{5,6,27,28,30} whereby Zn behaves like a ‘magnetic hole’ (a spinless site in an environment of strongly exchange-coupled spins) that strongly alters NN exchange correlations and disrupts superconductivity, whereas Ni retains a magnetic moment that barely perturbs the antiferromagnetic exchange correlations that facilitate superconductivity. Although the NMR and INS data^{24–31} are quite consistent with the magnetic component of such models, their predictions for local electronic phenomena at Ni and Zn can only now be tested for the first time. The STM data show that, despite their magnetic moments, scattering at Ni atoms is dominated by potential interactions. Furthermore, whereas Zn atoms locally destroy superconductivity within a 15 Å radius⁹, the magnetic Ni atoms coexist with unweakened superconductivity. All these phenomena are consistent with the above proposals.

The resilience of cuprate-oxide high- T_c superconductivity against what should be the destructive effects of a magnetic impurity atom, and its concomitant vulnerability to destruction by a ‘magnetic hole’, are remarkable. These atomic-scale phenomena are now (through a combination of NMR, μ SR and STM) coming into much clearer focus. They point towards a new approach to studying HTSC in which microscopic theories can be tested against an atomically resolved knowledge of impurity-state phenomena. □

Received 8 January; accepted 3 April 2001.

- Abrikosov, A. A. & Gorkov, L. P. Contribution to the theory of superconducting alloys with paramagnetic impurities. *Sov. Phys. JETP* **12**, 1243–1253 (1961).
- Monthoux, P., Balatsky, A. V. & Pines, D. Weak-coupling theory of high-temperature superconductivity in the antiferromagnetically correlated copper oxides. *Phys. Rev. B* **46**, 14803–14817 (1992).
- Moriya, T., Takehashi, Y. & Ueda, K. Antiferromagnetic spin fluctuations and superconductivity in two-dimensional metals—a possible model for high T_c oxides. *J. Phys. Soc. Jpn* **59**, 2905–2915 (1990).
- Bickers, N. E., Scalapino, D. J. & White, S. R. Conserving approximations for strongly correlated electron systems: Bethe-Salpeter equation and dynamics for the two-dimensional Hubbard model. *Phys. Rev. Lett.* **62**, 961–964 (1989).
- Monthoux, P. & Pines, D. Spin-fluctuation-induced superconductivity and normal-state properties of $\text{YBa}_2\text{Cu}_3\text{O}_7$. *Phys. Rev. B* **49**, 4261–4278 (1994).
- Pines, D. Understanding high temperature superconductors: progress and prospects. *Physica C* **282–287**, 273–278 (1997).
- Balatsky, A. V., Salkola, M. I. & Rosengren, A. Impurity-induced virtual bound states in d-wave superconductors. *Phys. Rev. B* **51**, 15547–15551 (1995).
- Salkola, M. I., Balatsky, A. V. & Schrieffer, J. R. Spectral properties of quasiparticle excitations induced by magnetic moments in superconductors. *Phys. Rev. B* **55**, 12648–12661 (1997).
- Pan, S. H. *et al.* Imaging the effects of individual zinc impurity atoms on superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **403**, 746–750 (2000).
- Mendels, P. *et al.* Macroscopic magnetic properties of Ni and Zn substituted $\text{YBa}_2\text{Cu}_3\text{O}_x$. *Physica C* **235/240**, 1595–1596 (1994).
- Salkola, M. I., Balatsky, A. V. & Scalapino, D. J. Theory of scanning tunneling microscopy probe of impurity states in a d-wave superconductor. *Phys. Rev. Lett.* **77**, 1841–1844 (1996).
- Flatté, M. E. & Byers, J. M. Impurity effects on quasiparticle c-axis planar tunneling and STM spectra in high- T_c cuprates. *Phys. Rev. Lett.* **80**, 4546–4549 (1998).
- Tsuchiura, H., Tanaka, Y., Ogata, M. & Kashiwaya, S. Local density of states around a magnetic impurity in high- T_c superconductors based on the t-J model. *Phys. Rev. Lett.* **84**, 3165–3168 (2000).
- Flatté, M. E. Quasiparticle resonant states as a probe of short-range electronic structure and Andreiev coherence. *Phys. Rev. B* **61**, R14920–14923 (2000).
- Haas, S. & Maki, K. Quasiparticle bound states around impurities in $d_{x^2-y^2}$ -wave superconductors. *Phys. Rev. Lett.* **85**, 2172–2175 (2000).
- Martin, I., Balatsky, A. V. & Zaanen, J. Impurity states and interlayer tunneling in high temperature superconductors. Preprint cond-mat/0012446 at <xxx.lanl.gov> (2000).
- Zhang, G.-M., Hu, H. & Yu, L. Marginal Fermi liquid resonance induced by quantum magnetic impurity in d-wave superconductors. *Phys. Rev. Lett.* **86**, 704–707 (2001).
- Yazdani, A., Howald, C. M., Lutz, C. P., Kapitulnik, A. & Eigler, D. M. Impurity-induced bound excitations on the surface of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Phys. Rev. Lett.* **83**, 176–179 (1999).
- Yazdani, A., Jones, B. A., Lutz, C. P., Crommie, M. F. & Eigler, D. M. Probing the local effects of magnetic impurities on superconductivity. *Science* **275**, 1767–1770 (1997).
- Flatté, M. E. & Byers, J. M. Local electronic structure of a single magnetic impurity in a superconductor. *Phys. Rev. Lett.* **78**, 3761–3764 (1997).
- Maeda, A., Yabe, T., Takebayashi, S., Hase, M. & Uchinokura, K. Substitution of 3d metals for Cu in $\text{Bi}_2(\text{Sr}_{0.8}\text{Ca}_{0.2})\text{Cu}_2\text{O}_x$. *Phys. Rev. B* **41**, 4112–4117 (1990).
- Kuo, Y. K. *et al.* Effect of magnetic and nonmagnetic impurities (Ni,Zn) substitution for Cu in $\text{Bi}_2(\text{SrCa})_{2+x}(\text{Cu}_{1-x}\text{M}_x)_{2+y}\text{O}_z$ whiskers. *Phys. Rev. B* **56**, 6201–6206 (1997).
- Bonn, D. A. *et al.* Comparison of the influence of Ni and Zn impurities on the electromagnetic properties of $\text{YBa}_2\text{Cu}_3\text{O}_{6.95}$. *Phys. Rev. B* **50**, 4051–4063 (1994).
- Ishida, K. *et al.* Cu NMR and NQR studies of impurities-doped $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)\text{O}_7$ (M=Zn and Ni). *J. Phys. Soc. Jpn* **62**, 2803–2818 (1993).

- Tokunaga, Y., Ishida, K., Kitaoka, Y. & Asayama, K. Novel relation between spin-fluctuation and superconductivity in Ni substituted high- T_c cuprate $\text{YBa}_2\text{Cu}_3\text{O}_7$: Cu NQR study. *Solid State Comm.* **103**, 43–47 (1997).
- Mahajan, A. V., Alloul, H., Collin, G. & Marucco, J. F. ^{89}Y NMR probe of Zn induced local moments in $\text{YBa}_2(\text{Cu}_{1-x}\text{Zn}_x)_3\text{O}_{6+x}$. *Phys. Rev. Lett.* **72**, 3100–3103 (1994).
- Bobroff, J. *et al.* Spinless impurities in high- T_c cuprates: Kondo-like behavior. *Phys. Rev. Lett.* **83**, 4381–4384 (1999).
- Bobroff, J. *et al.* Persistence of Li induced Kondo moments in the superconducting state of cuprates. *Phys. Rev. Lett.* **86**, 4116–4119 (2001).
- Williams, G. V. M., Tallon, J. L. & Dupree, R. NMR study of magnetic and non-magnetic impurities in $\text{YBa}_2\text{Cu}_3\text{O}_8$. *Phys. Rev. B* **61**, 4319–4325 (2000).
- Julien, M.-H. *et al.* ^{63}Cu NMR evidence for enhanced antiferromagnetic correlations around Zn impurities in $\text{YBa}_2\text{Cu}_3\text{O}_{6.7}$. *Phys. Rev. Lett.* **84**, 3422–3425 (2000).
- Sidis, Y. *et al.* Quantum impurities and the neutron resonance peak in $\text{YBa}_2\text{Cu}_3\text{O}_7$: Ni versus Zn. *Phys. Rev. Lett.* **84**, 5900–5903 (2000).
- Bernhard, C. *et al.* Suppression of the superconducting condensate in the high- T_c cuprates by Zn substitution and overdoping: Evidence for an unconventional pairing state. *Phys. Rev. Lett.* **77**, 2304–2307 (1996).

Acknowledgements

We acknowledge H. Alloul, P. W. Anderson, A. V. Balatsky, D. Bonn, M. Flatté, M. Franz, D.-H. Lee, K. Maki, I. Martin, P. Monthoux, A. Mourachkine, D. Pines, D. Rokhsar, S. Sachdev, D. J. Scalapino and A. Yazdani for conversations and communications, and J. E. Hoffman for help with data analysis. Support was from the Office of Naval Research, the Department of Energy through an LDRD from LBNL, the UCDRD Program, Grant-in-Aid for Scientific Research on Priority Area (Japan), a COE Grant from the Ministry of Education, Japan, the Miller Inst. for Basic Research (J.C.D.), and by the IBM Graduate Fellowship Program (K.M.L.).

Correspondence and requests for materials should be addressed to J.C.D. (e-mail: jcdavis@physics.berkeley.edu).

Liquid marbles

Pascal Aussillous & David Quéré

Laboratoire de Physique de la Matière Condensée, URA 792 du CNRS, Collège de France, 75231 Paris Cedex 05, France

The transport of a small amount of liquid on a solid is not a simple process, owing to the nature of the contact between the two phases. Setting a liquid droplet in motion requires non-negligible forces (because the contact-angle hysteresis generates a force opposing the motion¹), and often results in the deposition of liquid behind the drop. Different methods of levitation—electrostatic, electromagnetic, acoustic², or even simpler aerodynamic^{2,3} techniques—have been proposed to avoid this wetting problem, but all have proved to be rather cumbersome. Here we propose a simple alternative, which consists of encapsulating an aqueous liquid droplet with a hydrophobic powder. The resulting ‘liquid marbles’ are found to behave like a soft solid, and show dramatically reduced adhesion to a solid surface. As a result, motion can be generated using gravitational, electrical and magnetic fields. Moreover, because the viscous friction associated with motion is very small⁴, we can achieve quick displacements of the droplets without any leaks. All of these features are of potential benefit in microfluidic applications, and also permit the study of a drop in a non-wetting situation—an issue of renewed interest following the recent achievement of super-hydrophobic substrates⁵.

Liquid marbles are obtained by making a small amount of liquid (typically between 1 and 10 mm³) roll in a very hydrophobic powder (we used lycopodium grains of typical size 20 μm covered with fluorinated silanes). The grains spontaneously coat the drop, which can eventually be transferred onto other substrates. Figure 1 shows a marble of radius 1 mm, made of water and put on a glass plate (which would be wetted by water), where it is observed to adopt a spherical shape. Thanks to the monolayer of grains at the liquid–air interface, the wetting between the glass and the water is suppressed.

The marble can even be transferred onto a water pool, on which it is observed to float. Here we discuss the statics and dynamics of these marbles. By using mixtures of water and glycerol, the liquid viscosity was varied over a large range (a factor of 10^3), which allows us to describe the viscous and inertial dynamical regimes.

We first consider the size l of the contact between the marble and the solid. If there were a contact angle θ between the solid and the liquid (as usually occurs), l would be given by a simple geometric relation: $l = R \sin \theta$, where R is the radius of the droplet. Thus, l should tend to zero for non-wetting liquids ($\theta = 180^\circ$). However, the weight of the droplet needs to be considered⁴: the exact shape of the drop results from a balance between gravity (which favours a contact) and capillarity (which opposes it, because of the deformation it induces)⁴. If the centre of mass is lowered by a quantity δ , the difference in energy from a sphere tangent to the plane can be written dimensionally as: $\Delta E \approx \gamma \delta^2 - \rho g R^3 \delta$. Minimization of this expression yields δ , and thus, using the geometric relation $l^2 \approx \delta R$:

$$l \approx R^2 \kappa \quad (1)$$

where κ^{-1} is the capillary length ($\kappa^{-1} = \sqrt{\gamma/\rho g}$, where γ and ρ are, respectively, the liquid surface tension and the density). If the Laplace equation is solved numerically, the contact size can be plotted as a function of the drop radius. It is indeed found to be quadratic in R , with a numerical coefficient of the order of 0.8 (M. Perez, personal communication). For drops of initial radius R larger than the capillary length, gravity flattens the drop into a puddle. As the thickness of the puddle is fixed by gravity and surface tension, it must scale with the capillary length⁶. By invoking conservation of liquid volume, we can obtain the scaling law for the contact size: $l \approx R^{3/2} \kappa^{1/2}$.

A liquid marble indeed develops a contact with its substrate (Fig. 1). We measured the contact size l as a function of the drop radius, for two different powders (the lycopodium mentioned above, and a silanized silica powder of size ~ 100 nm) and two kinds of substrate (glass and Teflon plates). The results are displayed in Fig. 2a, where the data are found to collapse onto two curves. (1) For small droplets ($R < \kappa^{-1}$), the contact size indeed scales as R^2 , as predicted by equation (1), which proves that the contact is not a classical wetting one; the numerical coefficient is found to be 0.9, in close agreement with the numerical simulations. (2) Larger drops ($R > \kappa^{-1}$) form puddles, which yields the scaling behaviour in $R^{3/2}$.

We now consider the movement of such marbles on an inclined plane. We restrict our studies to the case of marbles formed from rather viscous liquids (viscosity $\eta > 200$ mPa s), in order to stress the small friction, related to the non-wetting, and the unusual laws of dissipation it generates. The motions are found to be stationary

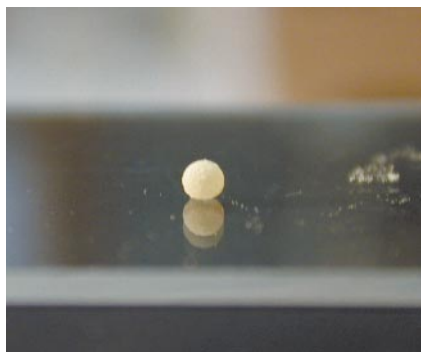


Figure 1 A liquid marble. Water is mixed with a very hydrophobic powder (silane-treated lycopodium grains of typical size 20 μ m). The powder spontaneously migrates to the water–air interface, and thus protects the water, which then behaves as if it were perfectly non-wetting when transferred on a glass plate. (Drop radius, 1 mm.)

after a short while (constant descent velocity, no liquid film remaining behind the drops). Thanks to the powder at the liquid–air interface, we can see that the marbles roll as they move.

The first series of experiments were performed with a small slope α (less than 10°). Usually, droplets would remain stationary in such a low field ($\sim 0.1g$), because of contact-angle hysteresis¹. In contrast, we found descent velocities that were rather large (of the order of 1 cm s^{-1}), even for very viscous fluids ($\sim 1,000$ mPa s). In Fig. 2b, this velocity is plotted as a function of the droplet radius. The velocity and drop size are respectively normalized by the puddle velocity V_0 (which should be independent of the puddle size) and by the capillary length κ^{-1} . Because of the normalization, all the data lie on the same curve. (1) For large drops (puddles), the velocity is indeed found to be independent of the drop size. It results from a balance between viscous friction and gravity. Because the puddle thickness is also independent of the drop size and fixed at a value of $2\kappa^{-1}$ (the maximum gravity pancake thickness, reached for an angle $\theta = 180^\circ$)⁶, the puddle velocity can be derived⁷: $V_0 \approx \gamma/\eta \sin \alpha$, which is drawn in Fig. 2b (the horizontal straight line), and found to fit the data. (2) For small droplets, the velocity increases when the drop size is decreased, which implies a very unusual friction law. This behaviour was recently predicted by Mahadevan and Pomeau⁴, and observed to apply qualitatively to droplets running down a so-called super-hydrophobic plate⁷. In the limit of small Reynolds numbers, Mahadevan and Pomeau⁴ propose that the droplet motion can be described as a superposition of a solid rotation (producing no dissipation) and a viscous friction localized in the contact zone.

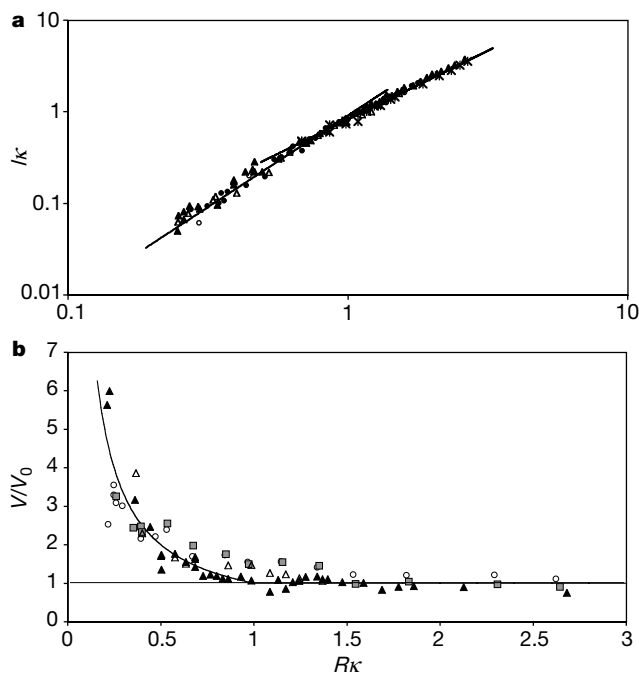


Figure 2 Contact of static liquid marbles, and their velocity on slightly tilted plates. **a**, The radius (l) of the contact between marble and substrate plotted against the radius (R) of the spherical drop. Both are normalized by the capillary length $\kappa^{-1} = (\gamma/\rho g)^{1/2}$. Lycopodium grains or silica were used as coating powders (represented by filled and open symbols, respectively). The liquid can be either water (triangles) or glycerol (circles). The marbles are deposited on glass, except in one case where glycerol was deposited on Teflon (stars). The straight lines represent slope 2 (equation (1)) and 1.5. **b**, Drop velocity (V) normalized by the puddle velocity (V_0) plotted against the drop radius (R) normalized by the capillary length (κ^{-1}), for different liquid viscosities, η , and slopes, α : open circles, $\eta = 270$ mPa s, $\alpha = 5^\circ$; grey squares, $\eta = 450$ mPa s, $\alpha = 5^\circ$; filled triangles, $\eta = 1,150$ mPa s, $\alpha = 4^\circ$; open triangles, $\eta = 1,150$ mPa s, $\alpha = 8^\circ$. The hyperbola is derived from equation (2), with 1 as a numerical constant.

Droplet deformation is small at low velocity, so that the size of this zone remains given by the static law (equation (1)). Typical velocity gradients in the drop are of the order of V/R , and the viscous force can be dimensionally written as: $f \approx \eta V/R^2$. Balancing the torque fR associated with this force by the gravity torque $\rho g R^4$ yields the marble velocity⁴:

$$V \approx V_0 \frac{\kappa^{-1}}{R} \quad (2)$$

As shown in Fig. 2b, such a hyperbolic law is found to fit the data fairly well, taking a numerical coefficient of the order of 1. This confirms that the hydrodynamics associated with these liquid marbles is very unusual, mainly because of the absence of a contact line. From a practical point of view, we stress that by dressing a drop with a hydrophobic powder, a weak field can drive the drop to a rather high speed (and higher if the drop is small).

The model of ref. 4 should be valid provided that the drop keeps a quasi-static shape, where inertial and viscous forces do not deform the spherical shape. This requires that both the Weber ($\rho V^2 R/\gamma$) and the capillary ($\eta V/\gamma$) numbers, which compare inertia and viscous force with surface tension, respectively, are less than unity. Together with the condition of small Reynolds number, this implies that the drop must be large enough ($R > \kappa^{-1} \sin \alpha$) and the liquid viscous enough ($\eta > (\gamma \rho \kappa^{-1} \sin \alpha)^{1/2}$). These conditions are fulfilled in the series of experiments displayed in Fig. 2b. We were interested in observing what happened when the influence of inertia was increased: this was examined by tilting the plate more (which generates larger velocities).

Figure 3 compares what happens when using a viscous marble on a small slope and on a greater slope ($\alpha = 4^\circ$ and $\alpha = 24^\circ$). Again, the velocity (scaled by the puddle velocity) is plotted versus the drop radius, and compared with the model of ref. 4 (equation (2)). A very large deviation is observed when the plate slope is increased, which reveals a regime of much smaller friction. The high velocities that are obtained (of the order of 1 m s^{-1}) induce strong deformations in the drop, as shown in Fig. 4, where successive snapshots of the drop (obtained with a high-speed camera) are displayed.

Two kinds of shape are observed, which can be both related to the action of centrifugal force. (1) A peanut shape, which has also been observed for drops rotating in space^{8,9}, and which has been discussed theoretically^{10,11}. (2) A more unusual deformation, which conserves the axial symmetry of the drop: as it accelerates, it gradually passes from a sphere to a disk, and eventually to a doughnut (toroidal) shape. The light intensity for a horizontal section crossing all the 'doughnut' centres is drawn in Fig. 4c. It may be seen that the 'holes' in the 'doughnuts' are not totally black. This means that the 'doughnut' is not fully open—opening could be delayed by the viscous drainage in the thin zone, around the centre of the doughnut, and implies the nucleation of a hole).

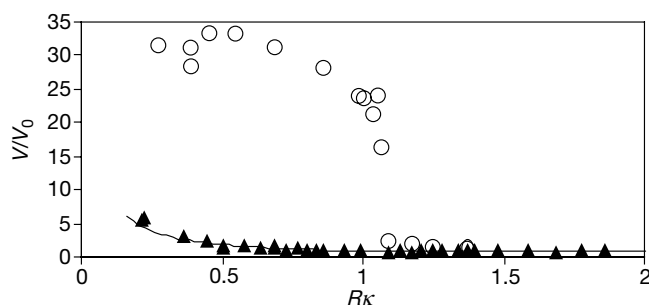


Figure 3 Comparison of the mobility of a liquid marble on two different slopes. The velocity V of viscous ($\eta = 1,150 \text{ mPa s}$) marbles (scaled by the puddle velocity V_0) is plotted against their radius R (scaled by the capillary length κ^{-1}), for $\alpha = 4^\circ$ (triangles) and $\alpha = 24^\circ$ (circles). The full line corresponds to equation (2).

A toroidal shape for a revolving drop has been proposed¹² and discussed¹⁰; but to the best of our knowledge such a shape has not been previously observed. (The formation of a torus has been reported when a very viscous liquid drop, held by a fibre inside another liquid of the same density, was rotated—which seems to be a different experiment¹³). The torus is claimed to be unstable compared with the peanut shape^{10,14}; its stability here seems to be promoted by the presence of an inclined plane below the marble. If this plane is removed (which may be done by putting an obstacle like a needle in the path of the drop, or more simply by observing what happens beyond the extremity of the plane), the axisymmetric shape evolves towards the peanut shape. This is illustrated in Fig. 4d. More generally, the torus seems to be metastable, and often evolves towards a peanut shape during its descent, usually because a small eccentricity in its shape makes it take off. A quantitative description of these different shapes, together with a comprehensive analysis of this low-friction regime, remains to be made.

Liquid marbles thus seem to be an interesting solution to the problem of driving small amounts of liquid on solid (or even liquid) substrates. In the present work we used gravity as a field to move the drop; but we have found that other weak fields (such as electric or magnetic fields) also drive liquid marbles (data not shown). Understanding the shape changes that we have observed will require the derivation of a friction law that takes shape evolution into account. We are at present examining the notable robustness of these marbles, which are able to resist a series of shocks (such as those observed in Fig. 4a). □

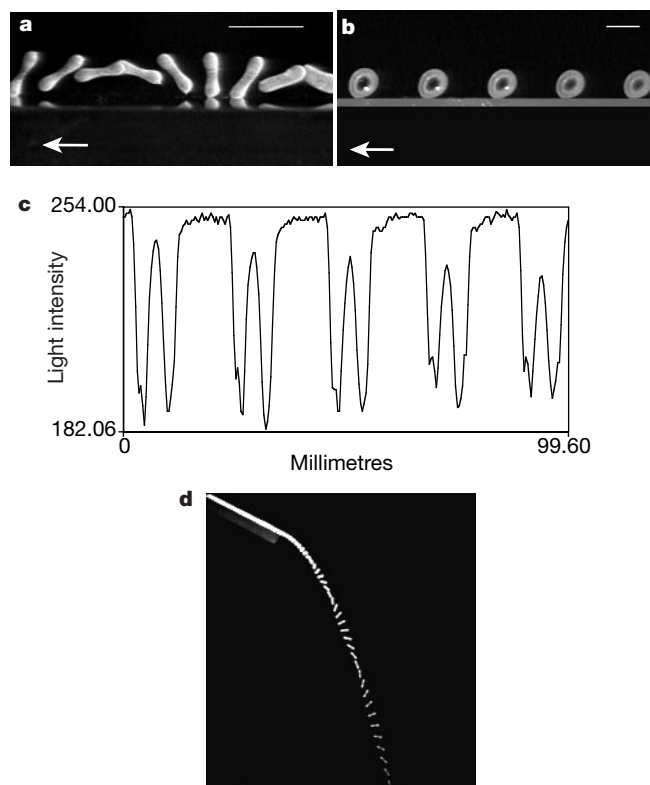


Figure 4 Different shapes taken by a liquid marble in the inertial regime. In pictures **a** and **b**, both the camera and the plane are tilted. Arrow, direction of the motion; scale bar, 1 cm. **a**, $\alpha = 34^\circ$, $R = 1.3 \text{ mm}$, interval between two pictures, 9 ms; **b**, $\alpha = 36^\circ$, $R = 2.5 \text{ mm}$, interval between two pictures, 23 ms. **c**, Light intensity for a horizontal section crossing all the 'doughnut' centres of **b**; 255 means black, and 0 white. **d**, Successive snapshots (interval between each, 6.7 ms) showing the evolution of an axisymmetric liquid marble as it leaves an inclined plane. The marble spontaneously transforms into a peanut shape.

Received 19 February; accepted 5 April 2000.

1. Dussan, V. E. B. & Chow, R. T. P. On the ability of drops or bubbles to stick to non-horizontal surfaces of solids. *J. Fluid Mech.* **137**, 1–29 (1983).
2. Frohn, A. & Roth, R. *Dynamics of Droplets* (Springer, Berlin, 2000).
3. Perez, M. *et al.* Oscillation of liquid drops under gravity: influence of shape on the resonance frequency. *Europhys. Lett.* **47**, 189–195 (1999).
4. Mahadevan, L. & Pomeau, Y. Rolling droplets. *Phys. Fluids* **11**, 2449–2453 (1999).
5. Onda, T., Shibuchi, S., Satoh, N. & Tsujii, K. Super water-repellent fractal surfaces. *Langmuir* **12**, 2125–2127 (1996).
6. Taylor, G. I. & Michael, D. H. On making holes in a sheet of fluid. *J. Fluid Mech.* **58**, 625–639 (1973).
7. Richard, D. & Quéré, D. Drops rolling on a tilted non-wettable solid. *Europhys. Lett.* **48**, 286–291 (1999).
8. Wang, T. G. *et al.* Bifurcation of rotating liquid drops: results from USML-1 experiments in space. *J. Fluid Mech.* **276**, 389–403 (1994).
9. Lee, C. P. *et al.* Equilibrium of liquid drops under the effects of rotation and acoustic flattening: results from USML-2 experiments in space. *J. Fluid Mech.* **354**, 43–67 (1998).
10. Brown, R. A. & Scriven, L. E. The shape and stability of rotating liquid drops. *Proc. R. Soc. Lond. A* **371**, 351–367 (1980).
11. Brown, R. A. & Scriven, L. E. New class of asymmetric shapes of rotating liquid drops. *Phys. Rev. Lett.* **45**, 180–183 (1980).
12. Rayleigh, Lord The equilibrium of revolving liquid under capillary force. *Phil. Mag.* **28**, 161–170 (1914).
13. Plateau, J. A. F. Experimental and theoretical researches on the figures of equilibrium of a liquid mass withdrawn from the action of gravity. *Annu. Rep. Board Regents Smithsonian. Inst.* 207–285 (1863).
14. Chandrasekhar, S. The stability of a rotating liquid drop. *Proc. R. Soc. Lond. A* **286**, 1–26 (1965).

Acknowledgements

We thank J. Bico and D. Richard for silanization of lycopodium grains and the achievement of a first series of liquid marbles, C. Clanet for help with high-speed pictures and for discussions, and P.-G. de Gennes for discussions and encouragement.

Correspondence and requests for materials should be addressed to D.Q. (e-mail: quere@ext.jussieu.fr).

Decreasing overflow from the Nordic seas into the Atlantic Ocean through the Faroe Bank channel since 1950

Bogi Hansen*, William R. Turrell† & Svein Østerhus‡

* Faroese Fisheries Laboratory, PO Box 3051, FO-110 Tórshavn, Faroe Islands

† FRS Marine Laboratory, PO Box 101, Aberdeen AB11 9DB, UK

‡ Bjerknes Centre for Climate Research and Geophysical Institute, N-5024 Bergen, Norway

The overflow of cold, dense water from the Nordic seas, across the Greenland–Scotland ridge¹ and into the Atlantic Ocean is the main source for the deep water of the North Atlantic Ocean². This flow also helps drive the inflow of warm, saline surface water into the Nordic seas¹. The Faroe Bank channel is the deepest path across the ridge, and the deep flow through this channel accounts for about one-third of the total overflow^{1,2}. Previous work has demonstrated that the overflow has become warmer and less saline^{3,4} over time. Here we show, using direct measurements and historical hydrographic data, that the volume flux of the Faroe Bank channel overflow has also decreased. Estimating the volume flux conservatively, we find a decrease by at least 20 per cent relative to 1950. If this reduction in deep flow from the Nordic seas is not compensated by increased flow from other sources, it implies a weakened global thermohaline circulation and reduced inflow of Atlantic water to the Nordic seas.

The Nordic seas and the Arctic Ocean are dominated by cold water below about 500 m depth—the typical sill depth of the Greenland–Scotland ridge (Fig. 1). The overflow of this water across the Greenland–Scotland ridge into the North Atlantic is the most important source for North Atlantic Deep Water²

(NADW), and any assessment of past or future variation in the global thermohaline circulation will depend critically on assessing the variation of this overflow. About half of the overflow passes through the Denmark Strait, while the rest crosses the ridge in the Iceland–Scotland region², mainly through the Faroese channels—consisting of the Faroe–Shetland channel leading to the Faroe Bank channel (Fig. 1b).

The Faroe Bank channel has a sill depth of around 840 m, and the deepest parts of the channel are constantly filled with cold overflow water flowing into the Atlantic with average velocities exceeding 1 m s^{-1} in the core (Fig. 2). Since November 1995, an upward-looking acoustic Doppler current profiler (ADCP) has been moored at the sill of the channel and has measured the velocity profile almost continuously. From these measurements, initiated in the ‘Nordic WOCE’ project, the flux of water colder than 3°C was estimated⁵ to be 1.9 Sv ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$), which is fairly consistent with previous estimates^{6,7}. This flux includes the overflow waters, but also some entrained Atlantic water. To avoid influence from the entrained water, we focus on the well defined overflow, colder than $+0.3^\circ\text{C}$ and with densities (σ_t) in excess of 28.0. From November 1995 to June 2000, the ADCP measurements indicate a weakening of this overflow flux of 2–4% per year (Fig. 3), which is significantly different from zero ($P < 0.001$).

Direct current measurements give the most reliable estimate of the overflow flux through the Faroe Bank channel, but for information on long-term variations we are restricted to more indirect methods. Historical hydrographic data may be used for this purpose if they can be related to the overflow on a theoretically sound basis. In the

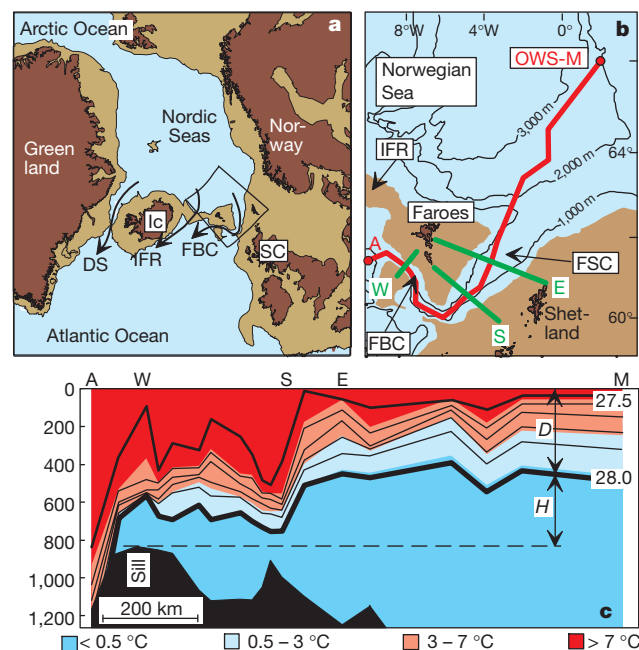


Figure 1 Topography and temperature change across the Greenland–Scotland ridge (light-brown areas are shallower than 500 m). **a**, Map of the Greenland–Iceland (Ic)–Scotland (Sc) region. Black arrows show overflow paths through the Denmark Strait (DS), over the Iceland–Faroe ridge (IFR), and through the Faroe Bank channel (FBC). **b**, Bottom topography of the study area (indicated by box in **a**) and location of a longitudinal (red line from A to OWS-M) and three transverse (green lines labelled W, S and E) standard hydrographic sections. Arrows indicate the Iceland–Faroe ridge (IFR), the Faroe–Shetland channel (FSC), and the Faroe Bank channel (FBC). **c**, Temperature (colour shading) and σ_t isopycnals (contours) along the longitudinal section, based on CTD (conductivity–temperature–depth) observations from RV *Magnus Heinason* in May–August 1991 and observations at OWS-M in the same period. The definition of the interface depth D and height H is illustrated.

steady state, the equations of motion may, to a first approximation, be integrated horizontally along the channel axis (x) and vertically (z) over the overflow layer to give:

$$\int \frac{1}{2} \rho U^2 dz - \int (\tau_{\text{top}} - \tau_{\text{bottom}}) dx = \int (p_u - p_s) dz = \Delta P \quad (1)$$

where U is the along-channel velocity at the sill, τ the frictional stress, and $(p_u - p_s)$ the difference between the upstream (p_u) and downstream, or sill (p_s), pressure. In the open Atlantic, outside the exit of the channel, the density above sill level referred to atmospheric pressure (ρ_0) is fairly constant. In the channel, the density structure is often approximated by a two-layer system, but at the upstream end it is better described as increasing linearly from ρ_0 in the surface to $\rho_0 + \Delta\rho$ (referred to atmospheric pressure) at the depth (D) of the interface, and then remaining constant through the overflow layer of height H (Fig. 1c). In this approximation, the integrated pressure difference is given by $\Delta P = (1/2) \Delta\rho g H (D + H)$.

The vertically integrated pressure difference between both ends of the channel (ΔP) is the driving force for the overflow, and because the frictional stress increases with velocity, equation (1) implies the intuitively obvious statement that the overflow flux is a monotonically increasing function of ΔP and hence also of H . This is the basis for the main conclusion of this Letter, as we infer a changing overflow flux (q) from an observed change in H . Quantitative relationships between q and H have been derived from analytical models^{6,8}, and generally imply that q is proportional to H raised to the power 3/2 or 2 depending on channel geometry. These models ignore friction, however, which is not permissible for the Faroe Bank channel^{7,9}. They also assume a pure two-layer structure. Using the more realistic expression for ΔP given above and assuming a narrow channel of rectangular cross-section, with constant velocity in the overflow layer and the interface height at the sill proportional to H , equation (1) implies:

$$q = \alpha H^n \quad (2)$$

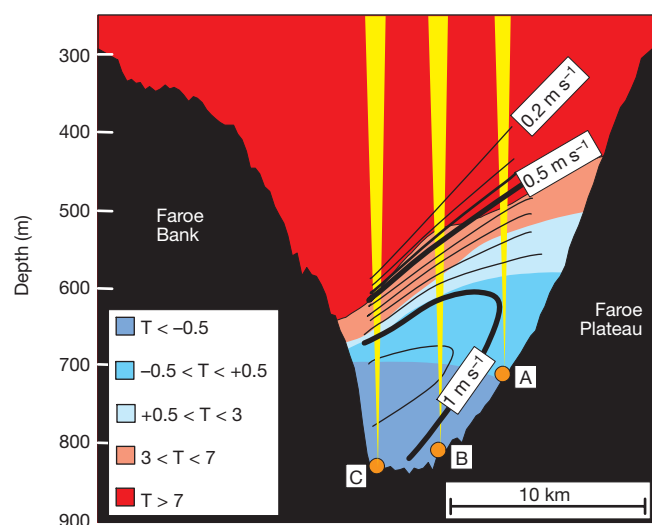


Figure 2 Section across the Faroe Bank channel (W in Fig. 1b) at the sill showing the temperature and velocity of the water below 250 m depth. Colour shading indicates the temperature distribution (in °C) observed on a cruise in September 1998. Contours show the average distribution of the along-channel velocity towards the northwest (positive into the paper) in the period July–September 1998 based on observations from three moored ADCPs (A, B and C) deployed during the VEINS project. Yellow cones illustrate sound beams. Assuming that the cross-sectional variation of velocity determined by this experiment is representative, volume flux below a given isotherm can be calculated from the central mooring alone.

where α is a constant, and $n = 1$ in the case where the first term (inertia) on the left hand side of equation (1) dominates, and $n = 3/2$ if the second term (friction) dominates and is quadratic in velocity. We use the constant-density level ($\sigma_t = 28.0$ isopycnal) to define the interface, and H is thus the height of this isopycnal above sill level (Fig. 1c).

Mesoscale activity deforms the isopycnals on timescales of days (Fig. 1c), which necessitates frequently sampled data. Such data are available at Ocean Weather Station M (OWS-M, Fig. 1b), where temperature and salinity have been observed since 1948 with several profiles a month¹⁰. Using these data, we established a time series of the depth of the $\sigma_t = 28.0$ isopycnal at OWS-M which exhibits a persistent deepening trend from 1950 to 1999 (Fig. 4). The annual depth increase was 1.05 ± 0.15 m (average $\pm$ standard error), which is statistically different from zero at a high level of significance ($P < 0.001$). This implies that the overflow also decreased over the 1950–99 period. Using equation (2), we find a decrease by 20% ($n = 1$) or by 26% ($n = 3/2$) relative to 1950.

A number of critical questions may be raised about the method and the associated assumptions. These are assessed in the Methods section, where we argue that they are unlikely to affect the basic conclusion of a weakened overflow flux. Fortunately, there is a way to test the assumption of a direct link between overflow flux through the Faroe Bank channel and interface height at OWS-M, because there is a period from November 1995 to January 1999 when simultaneous measurements exist of isopycnal depth at OWS-M and directly measured overflow flux in the Faroe Bank channel.

To compare these two sets of observations, we converted monthly interface depths at OWS-M to overflow flux using equation (2), smoothed them, and plotted them together with the flux computed from the ADCP measurements for the common period (Fig. 3). The two curves in Fig. 3 are not identical, but they show similar behaviour even at seasonal and interannual timescales as regards typical amplitudes and trends. This confirms that changes in the interface height at OWS-M reflect changes in overflow forcing, and supports our conclusion that the Faroe Bank channel overflow has decreased since 1950. This conclusion applies to water colder than 0.3°C , which accounts for about two-thirds of the total overflow flux (defined as water colder than 3°C). The remainder has densities ranging down to $\sigma_t = 27.8$, and this isopycnal has also deepened at OWS-M at similar rates since 1950. Thus, a decrease of the total Faroe Bank Channel overflow by about 0.5 Sv since 1950 is a conservative estimate.

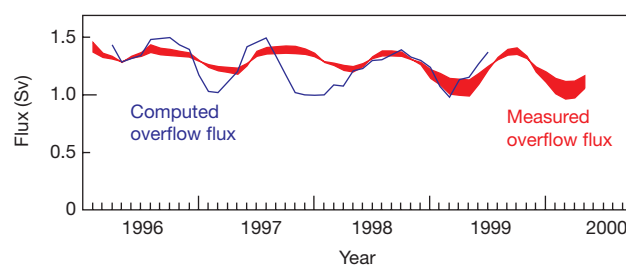


Figure 3 The flux of water below the 0.3°C isotherm through the Faroe Bank channel. Red curve, as measured by moored ADCPs; blue curve, as computed from OWS-M interface depth. Both series have been filtered using 5-month running means. The width of the red curve illustrates the uncertainty of the flux calculations due to a lack of information on the vertical isotherm movement in response to velocity changes⁵. The decreasing trend is estimated to be 2% or 4% per year, depending on the assumptions. The blue curve was computed from equation (2) with $n = 1$; α was determined by the average overflow flux and upstream interface height, and lagged by 8 months because independent hydrographic observations (see Methods) indicate that there is an 8-month delay in the interface height variations from OWS-M to the Faroe Bank channel.

Previous investigations^{3,4} have demonstrated a change in the composition of the Faroe Bank channel overflow that implies a reduction of the average density of the overflow water since the mid-1970s. Up to now, there was no observational evidence for changes in the overflow flux, but it has been argued³ that less-dense overflow water would entrain less ambient water in the Atlantic after leaving the channel. This would imply a reduced flux of the mixture of overflow water and entrained water feeding into the NADW. Our observed reduction of the overflow flux comes in addition to this. The Faroe Bank channel overflow has become less intensive and less dense, and both these effects have acted to reduce the supply of dense water to NADW.

The implications of our observed decrease of Faroe Bank channel flux depend on the behaviour of the other overflow sources. East of Iceland, the only source of sufficient magnitude to affect our conclusion is the overflow over the Iceland–Faroe ridge¹ (Fig. 1). The volume flux of this source can also be expected to decrease with a deepening interface in the southern Norwegian Sea. Thus, the total Iceland–Scotland overflow has probably decreased since 1950. In the Denmark Strait, analysis of hydrographic data since 1950 has indicated decadal variability in the overflow there¹¹, but this was based on the geostrophic method which is questionable in slope regions¹. Direct current measurements of the Denmark Strait overflow are more recent, but they show “surprisingly little evidence of variability over longer periods (months, seasons, years)”¹².

Thus, we cannot exclude the possibility of an increased Denmark Strait overflow that compensates for our observed decrease of Iceland–Scotland overflow, but we find no evidence for it. It has been shown that the formation of deep water by convection in the Greenland Sea has decreased since around 1970¹³, but that decrease might well have been compensated by a shift to more intensive formation of intermediate water¹. Unless there is an unobserved increase in flux from the Denmark Strait, our results suggest a reduction in the total formation of deep and intermediate waters north of the Greenland–Scotland ridge. This is consistent with predictions from global coupled atmosphere–ocean models as a consequence of an increase in anthropogenic greenhouse gases¹⁴.

The consequences for NADW production depend on entrainment as well as convection in the Labrador Sea, but, in any case, our results imply a reduced input of dense water from the Iceland–Scotland region and hence a reduced forcing of the meridional overturning from this overflow component.

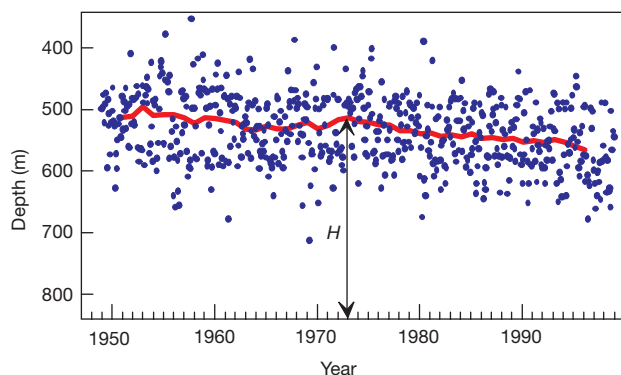


Figure 4 Depth of the $\sigma_t = 28.0$ isopycnal at OWS-M. Blue dots, monthly means. Red curve, 5-year running mean. Around 1970, the steady deepening of the interface seems to have been halted or even temporarily reversed at a time of intense deep-water formation in the Greenland Sea¹³. We note that during the decade 1990–99, the deepening of the $\sigma_t = 28.0$ isopycnal was about twice the typical decadal deepening for the whole period, which brings it closer to the trend measured by the ADCPs since 1995 (Fig. 3). H is the height of the isopycnal above sill level.

Our results also have implications for the Atlantic inflow to the Nordic seas. According to recent estimates¹, this inflow is mainly returned to the Atlantic in the overflows (75%) rather than in the surface flow (25%). This implies that the barotropic pressure gradient induced by the overflow is an important part of the driving force for the inflow¹. A reduced overflow can therefore be expected to lead to a reduced Atlantic inflow of a similar magnitude. The heat transported by the Atlantic inflow gives rise to the exceptionally mild climate of northwestern Europe¹⁵, and a reduction of the inflow will have a regional cooling effect that counteracts the global warming trend. As yet, direct flux measurements of the Atlantic inflow¹ are of too short duration to establish long-term trends, but reports of cooling and freshening in the Norwegian Sea¹⁶—and of decreasing air temperatures in some of the regions most affected by the inflow¹⁷—could perhaps be partly explained by a reduced inflow. □

Methods

Interface depth

The depth of the $\sigma_t = 28.0$ isopycnal at OWS-M was determined by linear interpolation between the standard observational depths: 400 m, 500 m, 600 m and 800 m. Thus, the depth interval between observations is twice the magnitude of the deepening trend. This, as well as mesoscale activity, argues that caution should be used in interpreting small isopycnal depth changes. We find, however, that the observed long-term deepening of 50 m is too large to be a statistical artefact. Spurious trends due to changing methodology, especially in salinity analysis, may also be discounted because they would have shown up also at deeper levels (beyond 1,500 m depth)—and they do not.

To establish the propagation of interface changes from OWS-M to the Faroe Bank channel, the timing of the seasonal maximum height of the $\sigma_t = 28.0$ isopycnal above sill depth was determined. At OWS-M, monthly average depths were fitted by a sinusoidal function with annual period added to a linear trend, and the maximal height was found to occur in January. In the Faroe channels, a similar analysis was performed using hydrographic observations from Scottish and Faroe research vessel cruises on three standard sections (Fig. 1b). There the maximal height was found to occur in June on section E and in August–September on sections S and W. This corresponds to an 8-month delay from OWS-M to the Faroe Bank channel.

Potential sources of error

It could be argued that the integrated pressure difference depends on other things rather than solely on the upstream interface height. One of these is the density in the Atlantic. We have examined a large number of stations from the 1950–99 period, obtained from the ICES (International Council for the Exploration of the Sea) Data Bank and the national Faroese data archive. Decadal averages of σ_t for each of the five decades ranged between 27.49 and 27.51 for the layer of depth 500–800 m. This verifies that the Atlantic waters just outside the exit of the channel have remained constant in density; but changes in the Norwegian Sea above the interface may also affect the horizontal pressure gradient at depth. The observations from the upper layers at OWS-M show that density in the upper layers has decreased slightly throughout most of the period, but this would tend to magnify the effect of a deepening interface on the overflow flux, not counteract it.

A difference in sea-level height between the Atlantic and the Norwegian Sea induces a barotropic pressure gradient. A systematic change in this could have offset the changing baroclinic pressure gradient implied by Fig. 4. Except for changing overflows, the only likely mechanism to induce a changing barotropic pressure gradient is the wind, and there certainly have been systematic changes in the wind—as documented by the well-known changes in the North Atlantic Oscillation (NAO) index¹⁸. These changes in the wind-forcing are, however, characterized by strengthening multi-decadal oscillations which are not seen in Fig. 4.

Using the interface height at OWS-M to reflect the overflow forcing presupposes that internal circulation in the Norwegian Sea does not change the interface depth between OWS-M and the channel entrance too much. Using high-quality data, available since 1994, we determined the average depth of the $\sigma_t = 28.0$ isopycnal from section E (Fig. 1b), close to the channel entrance, to be 562 ± 19 m (average $\pm$ standard error) in the period 1994–99. In the same period, the depth of this isopycnal at OWS-M was almost identical (565 ± 7 m). This argues that changes of the interface height between OWS-M and the channel entrance due to internal circulation are small compared to the 50-m change observed (Fig. 4).

Received 24 November 2000; accepted 19 April 2001.

- Hansen, B. & Østerhus, S. North Atlantic - Nordic Seas exchanges. *Prog. Oceanogr.* **45**, 109–208 (2000).
- Dickson, R. R. & Brown, J. The production of North Atlantic Deep Water: Sources, rates and pathways. *J. Geophys. Res.* **99**, 12319–12341 (1994).
- Turrell, W. R., Slessor, G., Adams, R. D., Payne, R. & Gillibrand, P. A. Decadal variability in the composition of Faroe Shetland Channel Bottom Water. *Deep-Sea Res.* **46**, 1–25 (1999).
- Hansen, B. & Kristiansen, R. Variations of the Faroe Bank Channel overflow. *Rit Fiskid.* **16**, 13–21 (1999).

5. Østerhus, S., Hansen, B., Kristiansen, R. & Lundberg, P. The Overflow through the Faroe Bank Channel. *Int. WOCE Newsl.* **35**, 35–37 (1999).
6. Borenäs, K. M. & Lundberg, P. A. On the deep-water flow through the Faroe Bank Channel. *J. Geophys. Res.* **93**, 1281–1292 (1988).
7. Saunders, P. M. Cold outflow from the Faroe Bank Channel. *J. Phys. Oceanogr.* **20**, 29–43 (1990).
8. Whitehead, J. A. Topographic control of oceanic flows in deep passages and straits. *Rev. Geophys.* **36**, 423–440 (1998).
9. Crease, J. The flow of Norwegian Sea Water through the Faroe Bank Channel. *Deep-Sea Res.* **12**, 143–150 (1965).
10. Østerhus, S. & Gammelsrød, T. The abyss of the Nordic Seas is warming. *J. Clim.* **12**, 3297–3304 (1999).
11. Bacon, S. Decadal variability in the outflow from the Nordic Seas to the deep Atlantic Ocean. *Nature* **394**, 871–874 (1998).
12. Dickson, R. R., Meincke, J., Vassie, I., Jungclauss, J. & Østerhus, S. Possible predictability in overflow from the Denmark Strait. *Nature* **397**, 243–246 (1999).
13. Clarke, R. A., Swift, J. H., Reid, J. L. & Koltermann, K. P. The formation of Greenland Sea deep water: Double diffusion or deep convection? *Deep-Sea Res.* **37**, 1385–1424 (1990).
14. Rahmstorf, S. Shifting seas in the greenhouse? *Nature* **399**, 523–524 (1999).
15. Rahmstorf, S. & Ganopolski, A. Long-term global warming scenarios computed with an efficient coupled climate model. *Clim. Change* **43**, 353–367 (1999).
16. Blindheim, J. et al. Upper layer cooling and freshening in the Norwegian Sea in relation to atmospheric forcing. *Deep-Sea Res.* **47**, 655–680 (2000).
17. Cappelen, J. & Laursen, E. V. *The Climate of the Faroe Islands* (Danish Meteorological Institute Technical Report 98-14, Copenhagen, 1998).
18. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science* **269**, 676–679 (1995).

Acknowledgements

The 'Nordic WOCE' project was funded by the Nordic Environmental Research Programme for 1993–97, and by Nordic national research councils. The 'VEINS' project was supported by the MAST IV Programme of the European Community.

Correspondence and requests for material should be addressed to B.H. (e-mail: bogihan@frs.fo).

Metamorphic core complex formation by density inversion and lower-crust extrusion

Fernando Martinez*, Andrew M. Goodliffe* & Brian Taylor†

* Hawaii Institute of Geophysics and Planetology, † Department of Geology and Geophysics, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA

Metamorphic core complexes are domal uplifts of metamorphic and plutonic rocks bounded by shear zones that separate them from unmetamorphosed cover rocks¹. Interpretations of how these features form are varied and controversial, and include models involving extension on low-angle normal faults², plutonic intrusions³ and flexural rotation of initially high-angle normal faults⁴. The D'Entrecasteaux islands of Papua New Guinea are actively forming metamorphic core complexes located within a continental rift that laterally evolves to sea-floor spreading⁵. The continental rifting is recent (since ~6 Myr ago)⁵, seismogenic⁶ and occurring at a rapid rate (~25 mm yr⁻¹)⁵. Here we present evidence—based on isostatic modelling, geological data and heat-flow measurements—that the D'Entrecasteaux core complexes accommodate extension through the vertical extrusion of ductile lower-crust material, driven by a crustal density inversion. Although buoyant extrusion is accentuated in this region by the geological structure present—which consists of dense ophiolite overlaying less-dense continental crust—this mechanism may be generally applicable to regions where thermal expansion lowers crustal density with depth.

The tectonic evolution of eastern Papua New Guinea created a crustal density inversion central to the proposed mechanism of core complex formation. The region of the palaeo-Papuan peninsula

west of 153° 20' E including the D'Entrecasteaux islands and the Goodenough basin (Fig. 1) originated in Palaeogene times as an intra-oceanic arc accommodating northerly directed subduction^{7,8}. The subducting plate included a continental fragment, the Papuan plateau, which had become detached from the margin of Australia by the opening of the Coral Sea basin⁹. When the Papuan plateau encountered the trench it was partially subducted⁹, and island arc ophiolite was obducted onto it^{7,10}. Northward subduction ceased and subduction later reversed¹¹. Southward subduction along the Trobriand trough led to arc magmatism throughout Miocene–Holocene times¹².

Sea-floor spreading in the easternmost Woodlark basin initiated by 6 Myr ago, and rifting around the D'Entrecasteaux islands had begun by Pliocene times¹² (Fig. 1). A seismic receiver function study¹³ indicates that the crust thins from 32–43 km beneath the Papuan peninsula to 20–26 km beneath the Goodenough basin and D'Entrecasteaux islands. On the D'Entrecasteaux islands, pressure–

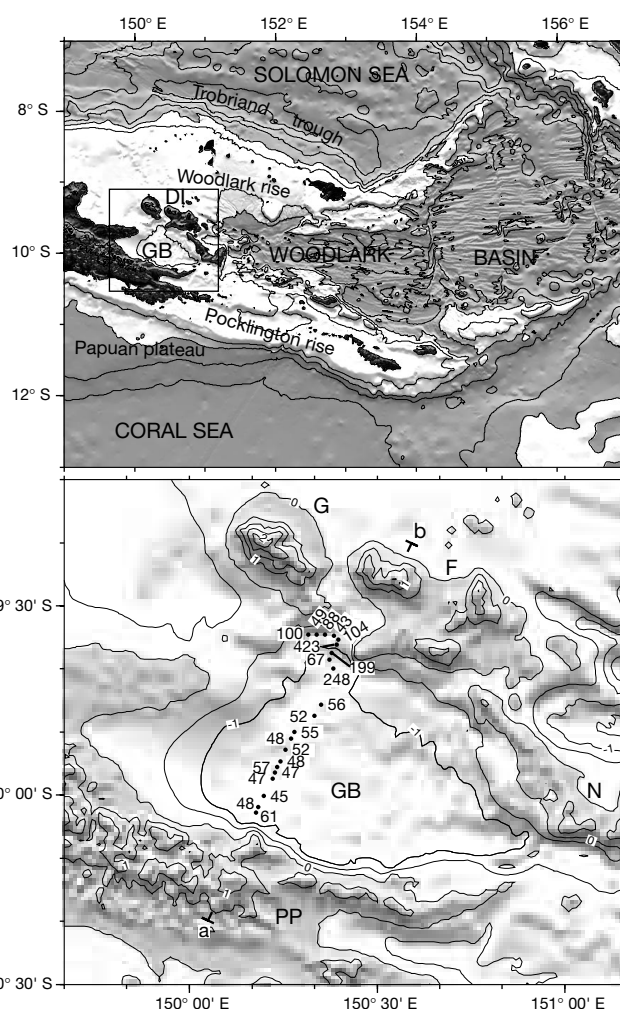


Figure 1 The eastern Papuan peninsula has been undergoing extension since at least 6 Myr ago. Top, sea-floor spreading in the oceanic Woodlark basin split the palaeo-peninsula to form the Woodlark and Pocklington rises as rifted margins. Extension decreases westwards. Continental rifting continues in the Goodenough basin and D'Entrecasteaux islands. Bottom, details of the study area indicated by the box in the top panel. Shown are the locations of the heat-flow measurements and their values (in mW m⁻²) uncorrected for sedimentation. Also shown are bathymetry and topography with contours at 500-m intervals annotated every kilometre. Abbreviations: DI, D'Entrecasteaux islands; GB, Goodenough basin; G, Goodenough island; F, Fergusson island; N, Normanby island; PP, Papuan peninsula. 'a–b' indicates modelled section shown in Fig. 4.

temperature–time studies of metamorphic minerals^{14,15} show that the cores of the islands were rapidly unroofed, rising to the surface from >10 kbar (~35 km depth) since about 4 Myr ago and cooling rapidly from initial temperatures of 570–730 °C. Although the islands lie within the seismogenic rift zone⁶ and within the region of crustal thinning¹³, they have undergone uplift rather than subsidence. Uplift continues to the present day¹⁵, forming topographic culminations reaching 1,500–2,500 m above sea level. In contrast, the Goodenough basin has subsided contemporaneously with the uplift of the islands, although seismic reflection studies¹⁶ indicate insufficient normal faulting to account for the degree of subsidence.

The rising metamorphic cores of the D'Entrecasteaux islands breached cover rocks including Palaeogene ophiolite and more recent volcanics and sediments^{8,17}. On the Papuan peninsula the ophiolite has an average thickness of ~15 km, including an 8–10-km-thick layer of basalt and gabbro and a 4–8-km-thick layer of harzburgite¹⁰. A similar crustal structure is supported regionally offshore: seismic refraction measurements¹⁸ from a line of shots across the northern coast of the Papuan peninsula (at Killerton station) find high seismic velocities (6.86–7.96 km s⁻¹) overlaying a deeper layer having lower velocities (6.5 km s⁻¹). ODP Leg 180¹⁹ recovered Palaeocene²⁰ gabbro and diabase in the area east of the D'Entrecasteaux islands, and oil-industry wells north of the islands also encountered Palaeocene ophiolite²¹. Bouguer gravity anomalies form relative lows over the islands²², which—together with seismic evidence precluding crustal roots¹³—indicates that the islands form low-density regions surrounded by higher-density crust. Thus, we infer that except over the islands where the cover rocks have been largely stripped off, the regional crustal structure consists of ophiolite overlaying metamorphic and plutonic continental crust (Fig. 2).

This crustal structure provides a mechanism for producing both uplift and subsidence within an area undergoing overall crustal thinning. Adopting typical density values for ophiolite (3,100 kg m⁻³) and continental crust (2,700 kg m⁻³) creates an

inverted density structure. Isostatically, if the 15-km-thick ophiolite layer were to be breached, the lower crustal layer could be extruded to a height of 2.2 km relative to the top of the ophiolite (Fig. 3a). Flow of lower crust from beneath the area would contribute to overall crustal thinning and subsidence without corresponding faulting, thinning and heating of the upper layer.

In order to examine these effects more quantitatively, we numerically modelled the thermal structure and isostatically predicted elevations for various forms of extension, and compared them to heat-flow measurements and topography. In a young rift undergoing rapid extension, a significant heat-flow anomaly can be produced that is sensitive to the pattern of lithospheric and crustal thinning. We made heat-flow measurements across Goodenough basin, from offshore of the Papuan peninsula to near Fergusson island (Fig. 1; see also Methods section). South of 9° 41' S the heat-flow measurements average 51 mW m⁻² and have low variability (45–61 mW m⁻²). To the north, near the island core complexes, the measurements have a distinctly higher average (147 mW m⁻²) and greater variability (43–423 mW m⁻²). Accounting for topographic effects and sedimentation following standard methodologies^{23,24} increases average values south of 9° 41' S to 73 mW m⁻² and to the north near the islands to 190 mW m⁻². Small Quaternary volcanic centres occur on the D'Entrecasteaux islands⁸, and sea-floor side-scan sonar imagery and multi-channel seismic profiles^{16,25} suggest small intrusions in the offshore area of the northern heat-flow measurements. These volcanic features and the higher and more variable heat flow imply a significantly higher thermal regime near the islands compared to the basin to the south.

An estimate of the crustal and lithospheric thickness and heat flow before the formation of the islands can be made by adopting a thickness of 40 km and a temperature of 700 °C for the base of the crust before 4 Myr ago—these values are consistent with the metamorphic pressure–temperature–time studies on the islands^{14,15}, and with seismic estimates¹³ of the crustal thickness today on the Papuan peninsula. These values give a heat flow of 56 mW m⁻² and a lithospheric thickness of ~74 km, assuming a

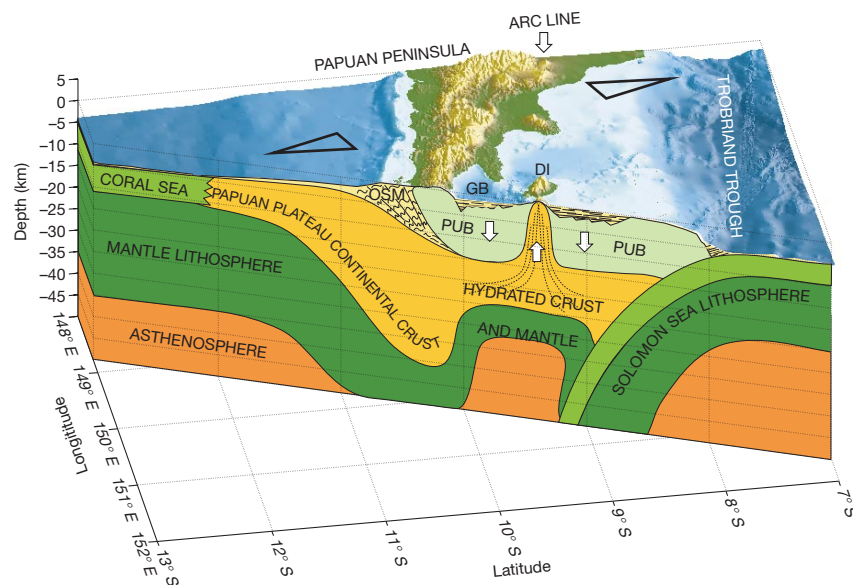


Figure 2 Schematic interpretation of the present-day crustal structure along the heat-flow profile extended from the Coral to Solomon seas. A Palaeogene arc–continent collision emplaced the Papuan ultramafic belt (PUB), an ophiolite sequence, over the Papuan plateau, a continental fragment of Australia and cover sediments that now form the Owen Stanley metamorphic (OSM) belt on the peninsula. Arc reversal led to the Solomon Sea subducting southward along the Trobriand trough. The current overall extension direction is indicated by triangles. The D'Entrecasteaux islands metamorphic core complexes

formed when the PUB was breached and pulled apart by the regional extension, allowing the lower ductile continental crust to be extruded by the weight of the overlying PUB. Flow of the lower crust led to its preferential thinning and subsidence in the Goodenough basin (GB). Heat associated with the arc line and hydration of the lower crust and mantle by fluids from the subducting Solomon Sea plate locally weakened this area, permitting the lower crust to flow. Moho relief beneath the cold forearc and Papuan peninsula outside of this area is maintained by the stronger crust and mantle there.

linear geotherm, lithospheric conductivity of $3.2 \text{ W m}^{-1} \text{ } ^\circ\text{C}^{-1}$ and an asthenospheric temperature of $1,300^\circ\text{C}$. We used two numerical techniques to model the two-dimensional, time-dependent thermal evolution and crustal and lithospheric thinning since 4 Myr ago. In the first technique^{26,27}, values of lower lithospheric extension (β) and upper lithospheric extension (δ) are specified separately. In the second, a finite-difference method²⁸ is used to calculate the thermal structure for a specified geometry of pure shear extension. Thermal diffusivity is $1 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$ in both cases. We first model uniform pure shear lithospheric extension by a factor of 1.6, which thins the crust from 40 km beneath the peninsula to 25 km beneath the basin and islands, and predicts heat flow of $\sim 90 \text{ mW m}^{-2}$ in the basin. The significantly lower measured and corrected heat flow in most of the basin thus argues against uniform pure shear extension. Assuming the same starting conditions, but using a value of $\beta = 2.5$ for the crust and lithosphere beneath the 15-km-thick ophiolite layer, and using $\delta = 1$ (no extension) for the ophiolite layer, also results in an overall crustal thinning from 40 km to 25 km. In this case, however, a lower heat flow in the basin interior of 74 mW m^{-2} results, due primarily to a time lag for heat to conduct across the upper unthinned layer. This lower value is more consistent with the

sedimentation-corrected heat-flow values. An aspect of two-layer stretching is that it does not conserve volume²⁶, and in our implementation of this method the use of symmetry about the D'Entrecasteaux islands augments this artefact. The lack of two-dimensional across-axis volume conservation in problems involving lower crustal flow may, however, reflect actual three-dimensional flow. Neither uniform pure shear of the entire crust nor two-layer extension by itself, however, predicts the uplift of the islands.

We investigate the formation of the islands using the second technique. We model the splitting and separation of the upper 25 km of crust within a narrow zone (4 km) of pure shear deformation overlaying a 700°C isothermal ductile lower crust which can flow into the space created. The initial temperature profile in the upper 15 km is varied so that the area outside of the zone of pure shear extension matches the two-layer extension result at 4 Myr. Because the zone of stretching is narrow, the upper dense ophiolite layer is quickly thinned and replaced by hot, less-dense, lower continental crust. Rapid isostatic uplift of the islands results primarily from the compositional density contrast of the layers, but there is also a smaller component of thermal uplift (Fig. 3b)

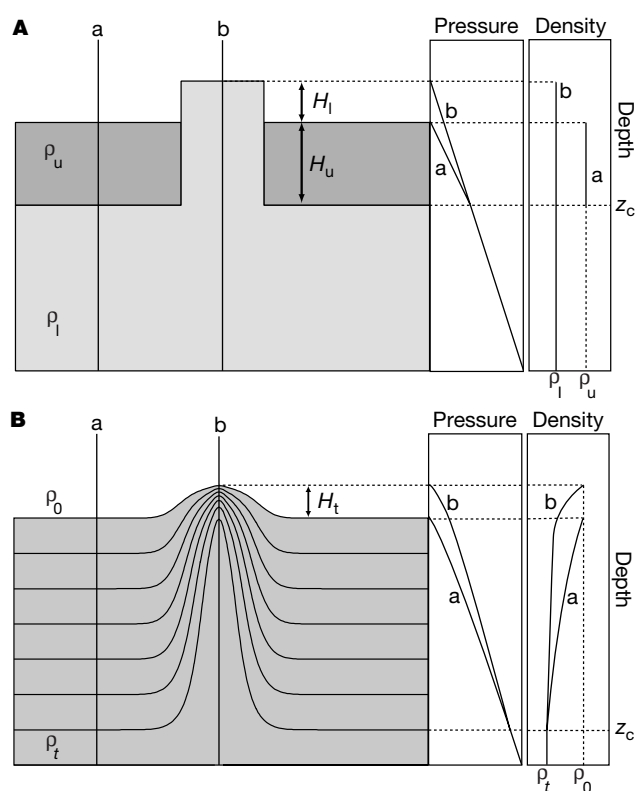


Figure 3 Isostatic model of lower crustal flow and extrusion driven by density inversion and localization of extension. **A**, Compositionally stratified crust in which an upper layer with density ρ_u greater than that of a lower layer ρ_l is stretched, forming a break in the upper layer. The lower layer is extruded and rises to a height H_l such that the pressure of the two columns, a and b, at z_c , the compensation depth, are equal: $g\rho_l(H_l + H_u) = g\rho_u H_u$, where g is the acceleration of gravity. Panels at right show the variation in pressure and density with depth for the two columns, a and b. **B**, Constant-composition crust in which the density at temperature t , ρ_t , is related to the density at 0°C , ρ_0 , by thermal expansion coefficient, α , following $\rho_t = \rho_0(1 - \alpha t)$. Column a has a background linear increase in temperature with depth. Column b is located at the centre of the narrow area undergoing pure shear extension. The low-density, hot lower layer has been rapidly extruded. As column b has lower density than column a above the compensation depth (z_c) it rises a height H_t to achieve isostatic equilibrium.

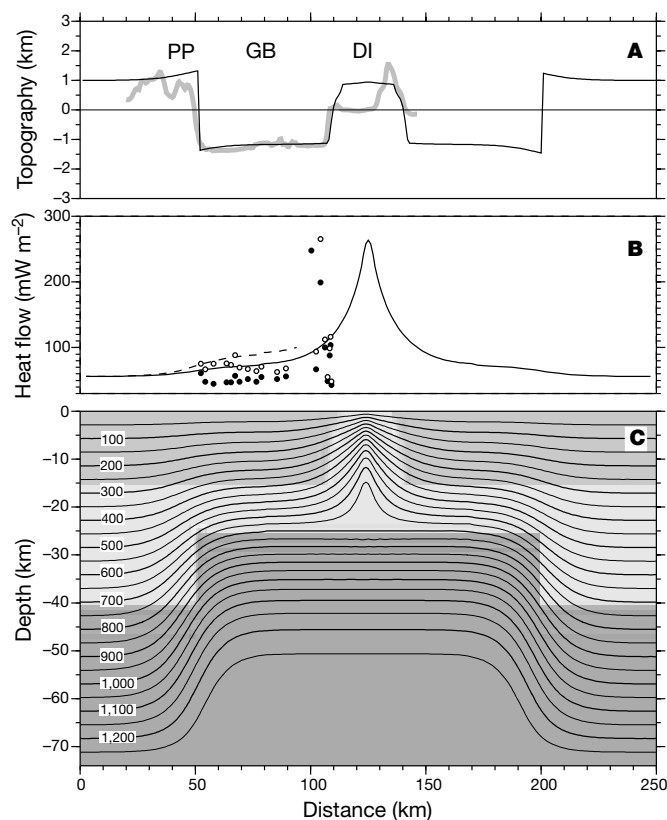


Figure 4 Two-dimensional, time-dependent thermal calculations combining the effects of two extension calculations. **A**, Predicted topography (solid line) and observed topography (grey line) along the heat flow traverse in the Goodenough basin (GB) extended to the Papuan peninsula (PP) and D'Entrecasteaux islands (DI) (a–b in Fig. 1b). **B**, Predicted heat flow (solid line), measured heat flow (filled circles) and heat flow corrected for sedimentation and topographic effects (open circles). Dashed line shows the higher heat flow predicted by the uniform pure shear extension calculation relative to the two-layer extension. **C**, Composite crustal geometry showing the upper ophiolite layer (intermediate grey), lower continental crust (light grey), mantle (dark grey) and thermal structure after 4 Myr of extension. Extension geometries and other parameters are discussed in the text. Finite difference calculations for the core complex formation were merged between horizontal distance 80–170 km and depth 0–25 km. The area of high topography between 200 and 250 km is an artefact of model symmetry centred on the D'Entrecasteaux islands, and is not intended to model the forearc area.

which increases over the 4 Myr of extension due to heat advecting into the crustal section undergoing stretching.

We combined the thermal structures resulting from the two-layer and finite-difference thermal calculations, along a section where they predicted essentially the same crustal temperatures and heat flow (within 1 mW m^{-2}). This thermal structure was then combined with the predicted crustal compositional geometry to calculate densities and the isostatically balanced topography. Densities (at 0°C) of 3,100, 2,700, and $3,300 \text{ kg m}^{-3}$ were used for the ophiolite, lower continental crust, and mantle, respectively, with a uniform thermal expansion coefficient of $3.4 \times 10^{-5} \text{ }^\circ\text{C}^{-1}$. The results (Fig. 4) show that the seismically determined overall crustal thinning beneath the basin and islands can be achieved by thinning the lower crustal layer, and that this is consistent with basin subsidence and heat flow. Simultaneous breaching of the upper layer and buoyant extrusion of lower crust can explain the elevation of the core complexes relative to the basin, and also produces a distinct heat flow 'high' in their vicinity in broad agreement with thermal measurements.

These calculations are intended to illustrate general features of the proposed mechanisms of core complex formation by lower crustal flow and buoyant extrusion rather than constitute a rigorous modelling of the details of this area. We summarize below these features and their more general implications for continental rifting.

(1) Geological and geophysical observations indicate that the regional crustal structure surrounding the D'Entrecasteaux islands consists of ophiolite overlaying less-dense continental crust, creating a two-layer inverted crustal density profile.

(2) Overall crustal thinning is seismically observed in the area of the Goodenough basin and D'Entrecasteaux islands and is necessary to explain the subsidence of the basin, but uniform crustal thinning is inconsistent with heat-flow values, estimates of extension by faulting, and relative uplift of the islands.

(3) Low heat-flow measurements in the basin interior relative to model predictions for uniform crustal thinning suggest that preferential thinning of the lower crust relative to the upper crust occurs across the basin. We infer that this thinning occurs both by stretching and by crustal flow into the core complexes.

(4) Formation of the island core complexes is explained by buoyant extrusion of ductile lower crust enabled by splitting and pulling apart of the ophiolite layer, locally focusing the regional extension. The buoyancy of the lower crust alone would not probably be sufficient to breach the stronger upper layer.

(5) Because the core complex emplacement accommodates the locally focused extension it does not generate compression, as in gravity sliding models²⁹, nor does it generate purely radial shearing patterns implied by models of forceful plutonic emplacement³⁰.

(6) A narrow zone of extension can produce very rapid vertical advection of the lower crust, and account for the nearly isothermal ascent of material followed by rapid cooling inferred from the metamorphic pressure-temperature-time studies. The rapid ascent of lower crust also produces a local heat-flow 'high' in broad agreement with the high thermal regime near the islands.

(7) Although lower crustal flow, implying a weak lower crust, is proposed to occur beneath the basin and islands, significant Moho relief, implying a strong lower crust, is nevertheless maintained between the peninsula and basin. We suggest that this dual behaviour is related to the subduction of the Solomon Sea slab (Fig. 2) that locally introduces heat associated with the arc line and hydrates the mantle and lower crust beneath the basin and islands—this makes these areas weaker than those under the peninsula and cold forearc region.

(8) The mechanism of buoyant lower crustal extrusion may occur in other areas of extension and core complex formation where earlier obduction or thrusting has emplaced layers of greater density over less dense ones (Fig. 3a). It may also occur in areas of more uniform crustal composition where increasing temperatures with

depth lowers crustal density by thermal expansion (Fig. 3b). In this case a cooler and more brittle surface layer may locally rupture in extension, allowing the deeper, hotter, less dense, and plastic layer to be rapidly extruded. □

Methods

Heat-flow data acquisition and reduction

Heat-flow data were acquired using the 'violin bow' apparatus of the Pacific Geoscience Centre, Geological Survey of Canada; this apparatus consists of a 3-m-long sensor string enclosing 11 thermistors mounted outboard of a strength member. *In situ* conductivities were determined at each measurement site by discharging a calibrated heat pulse into the sediments. Water column temperatures across the basin were constant below $\sim 600 \text{ m}$ and plots of temperature versus thermal resistance in the sediments were linear, indicating no disturbances due to bottom-water temperature variation or advection in the sediments. Data were corrected for sedimentation using a six-channel seismic profile to estimate sediment thickness to acoustic basement at each heat-flow measurement site. A limit on the plausible sedimentation effects was estimated using an analytical one-dimensional solution²⁴, assuming that the observed thickness at each site was deposited in 1 Myr and using a sediment thermal diffusivity of $0.421 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$. Topographic corrections were estimated using a two-dimensional relaxation method²⁵ and the seismic profile to estimate relief on the sea floor and acoustic basement—assuming thermal conductivities of 1.2 and $2.3 \text{ W m}^{-1} \text{ K}^{-1}$ for sediments and acoustic basement, respectively.

Received 22 January; accepted 18 April 2001.

1. Coney, P. J. in *Cordilleran Metamorphic Core Complexes* (eds Crittenden, M. D., Coney, P. J. & Davis, G. H.) 7–31 (The Geological Society of America, Boulder, 1980).
2. Wernicke, B. Low-angle normal faults in the Basin and Range Province: nappe tectonics in an extending orogen. *Nature* **291**, 645–648 (1981).
3. Lister, G. & Baldwin, S. Plutonism and the origin of metamorphic core complexes. *Geology* **21**, 607–610 (1993).
4. Buck, W. R. Flexural rotation of normal faults. *Tectonics* **7**, 959–973 (1988).
5. Taylor, B., Goodliffe, A. M. & Martinez, F. How continents break up: insights from Papua New Guinea. *J. Geophys. Res.* **104**, 7497–7512 (1999).
6. Abers, G. A., Mutter, C. Z. & Fang, J. Shallow dips of normal faults during rapid extension: Earthquakes in the Woodlark-D'Entrecasteaux rift system, Papua New Guinea. *J. Geophys. Res.* **102**, 15301–15318 (1997).
7. Rogerson, R., Hilyard, D., Francis, G. & Finlayson, E. The foreland thrust belt of Papua New Guinea. *Proc. Pacif. Rim Congr.* **87**, 579–583 (1987).
8. Davies, H. L. & Warren, R. G. Origin of eclogite-bearing, domed, layered metamorphic complexes ("core complexes") in the D'Entrecasteaux Islands, Papua New Guinea. *Tectonics* **7**, 1–21 (1988).
9. Weissel, J. K. & Watts, A. B. Tectonic evolution of the Coral Sea basin. *J. Geophys. Res.* **84**, 4572–4582 (1979).
10. Davies, H. L. Peridotite-gabbro-basalt complex in eastern Papua: An overthrust plate of oceanic mantle and crust. (Bulletin No. 128, Bureau of Mineral Resources, Geology and Geophysics, Canberra, 1971).
11. Davies, H. L., Symonds, P. A. & Ripper, I. D. Structure and evolution of the southern Solomon Sea region. *BMR J. Aust. Geol. Geophys.* **9**, 49–68 (1984).
12. Taylor, B. Background and regional setting. *Proc. ODP Init. Rep.* [CD-ROM] (eds Taylor, B. et al.) **180**, 1–20 (Ocean Drilling Program, Texas A & M Univ., College Station, 1999).
13. Ferris, A. et al. Active continental extension and metamorphic core complexes: A PASSCAL seismic deployment in the D'Entrecasteaux Islands, eastern Papua New Guinea. *Eos* **81**, F881 (2000).
14. Baldwin, S. L., Lister, G. S., Hill, E. J., Foster, D. A. & McDougall, I. Thermochronologic constraints on the tectonic evolution of an active metamorphic core complex, D'Entrecasteaux Islands, Papua New Guinea. *Tectonics* **12**, 611–628 (1993).
15. Hill, E. J., Baldwin, S. L. & Lister, G. S. Unroofing of active metamorphic core complexes in the D'Entrecasteaux Islands, Papua New Guinea. *Geology* **20**, 907–910 (1992).
16. Fang, J. *Styles and Distribution of Continental Extension Derived from the Rift Basins of Eastern Papua New Guinea*. Thesis, Columbia Univ. (2000).
17. Davies, H. L. & Warren, R. G. Eclogites of the D'Entrecasteaux Islands. *Contrib. Mineral. Petrol.* **112**, 463–474 (1992).
18. Finlayson, D. M., Drummond, B. J., Collins, C. D. N. & Connelly, J. B. in *Volcanism in Australasia* (ed. Johnson, R. W.) 259–274 (Elsevier, Amsterdam, 1976).
19. Shipboard Scientific Party. Leg 180 Summary. *Init. Rep. ODP* (eds Taylor, B., et al.) 1–77 (Ocean Drilling Program, Texas A & M Univ., College Station, 1999).
20. Baldwin, S. L., Monteleone, B., Hill, E. J., Ireland, T. R. & Fitzgerald, P. G. Continental extension in the western Woodlark basin, Papua New Guinea. *Eos* **81**, F1307 (2000).
21. Francis, G., Lock, J. & Okuda, Y. Seismic stratigraphy and structure of the area to the southeast of the Trobriand Platform. *Geo-Mar. Lett.* **7**, 121–128 (1987).
22. Milsom, J. The gravity field of the Papuan Peninsula. *Geol. Mijnb.* **52**, 13–20 (1973).
23. Sclater, J. G., Jones, E. J. W. & Miller, S. P. The relationship of heat flow, bottom topography and basement relief in Peake and Freen Deeps, northeast Atlantic. *Tectonophysics* **10**, 283–300 (1970).
24. Langseth, M. G., Hobart, M. A. & Horai, K. Heat flow in the Bering Sea. *J. Geophys. Res.* **85**, 3740–3750 (1980).
25. Taylor, B., Goodliffe, A., Martinez, F. & Hey, R. N. Continental rifting and initial seafloor spreading in the Woodlark basin. *Nature* **374**, 534–537 (1995).
26. Royden, L. & Keen, C. E. Rifting processes and thermal evolution of the continental margin of eastern Canada determined from subsidence curves. *Earth Planet. Sci. Lett.* **51**, 343–361 (1980).
27. Steckler, M. S. *The Thermal and Mechanical Evolution of Atlantic-Type Continental Margins*. Thesis, Columbia Univ. (1981).
28. Buck, W. R., Martinez, F., Steckler, M. S. & Cochran, J. R. Thermal consequences of lithospheric extension: Pure and simple. *Tectonics* **7**, 213–234 (1988).

29. Seager, W. R. Low-angle gravity glide structures in the northern Virgin Mountains, Nevada and Arizona. *Geol. Soc. Am. Bull.* **81**, 1517–1538 (1970).
30. Ollier, C. D. & Pain, C. F. Active rising surficial gneiss domes in Papua New Guinea. *J. Geol. Soc. Aust.* **27**, 33–44 (1980).

Acknowledgements

We thank E. Davis for providing the heat flow probes used in this study; T. Lewis and A. Taylor for their help and expertise in acquiring and reducing heat flow measurements at sea; the Captain and crew of the RV *Maurice Ewing* for their efforts which led to a successful cruise; and R. Buck and S. Baldwin for comments and suggestions which improved the paper. This work was supported by the US National Science Foundation.

Correspondence and requests for materials should be addressed to E.M.
(e-mail: martinez@soest.hawaii.edu).

The elastic constants of MgSiO_3 perovskite at pressures and temperatures of the Earth's mantle

Artem R. Oganov, John P. Brodholt & G. David Price

Department of Geological Sciences, University College London, Gower Street, London WC1E 6BT, UK

The temperature anomalies in the Earth's mantle associated with thermal convection¹ can be inferred from seismic tomography, provided that the elastic properties of mantle minerals are known as a function of temperature at mantle pressures. At present, however, such information is difficult to obtain directly through laboratory experiments. We have therefore taken advantage of recent advances in computer technology, and have performed finite-temperature *ab initio* molecular dynamics simulations^{2,3} of the elastic properties of MgSiO_3 perovskite, the major mineral of the lower mantle, at relevant thermodynamic conditions. When combined with the results from tomographic images of the mantle, our results indicate that the lower mantle is either significantly anelastic⁴ or compositionally heterogeneous on large scales⁵. We found the temperature contrast between the coldest and hottest regions of the mantle, at a given depth, to be about 800 K at 1,000 km, 1,500 K at 2,000 km, and possibly over 2,000 K at the core–mantle boundary.

Ab initio molecular dynamics, pioneered by Car and Parrinello², has made a substantial impact on modern planetary and materials sciences^{6–8}. Advantages of this approach—which is computationally expensive—include the explicit account of the quantum nature of electronic structure and interatomic interactions, and the description of the intrinsic anharmonicity of lattice vibrations. This is the method of choice for simulating high-temperature phenomena that are significantly anharmonic, such as melting^{8,9}, ionic conductivity, displacive phase transitions, thermal expansion¹⁰, and elastic properties. Calculations of elastic constants of materials using *ab initio* molecular dynamics can be done by one of two routes: (1) from fluctuations of stress or strain, or (2) from stress–strain relations, that is, from Hooke's law. Both routes require enormous computational efforts; very long runs are required by the first method, and in the second method, it is necessary to perform several independent simulations for different lattice strains that include the nonlinearity of stress–strain relations. Recent advances in computer technology and increased accessibility of supercomputers have now made such simulations possible.

Orthorhombic MgSiO_3 perovskite is an ideal subject for *ab initio* molecular dynamics simulations. When containing some Fe, it is the most abundant mineral in the Earth, comprising 60–100 vol.% of

the lower mantle, and dominating many of its properties. The lower mantle comprises over 50% of the Earth's volume; extending between the depths of 670 km and 2,891 km, it is characterized by very high pressures (24–136 GPa) and temperatures, roughly between 2,000 K and 3,000 K, possibly rising to about 4,000 K (ref. 11). At these temperatures atomic motion is essentially classical and possibly significantly anharmonic, making it appropriate to use the molecular dynamics approach. A hypothetical cubic phase with superionic conductivity has been proposed as an explanation of the observed high electrical conductivity of the lower mantle^{12,13}; molecular dynamics is the only currently viable way to simulate fast ionic conduction phenomena in solids. Our previous work¹⁴ considered 'static' elastic constants (that is, the thermal motion of atoms was neglected), and used *ab initio* molecular dynamics to study the thermal expansivity and the thermal equation of state of MgSiO_3 perovskite—which was found to stay orthorhombic at lower-mantle temperatures. Here we perform *ab initio* molecular dynamics simulations of elastic constants of MgSiO_3 perovskite at lower-mantle conditions, taking full account of temperature. At this stage we consider only pure MgSiO_3 perovskite. The effects of the moderate Fe content, while potentially important for certain properties (such as shear modulus), are expected to be negligible for others—especially for the thermal expansion, bulk modulus, and derivatives (with respect to pressure and temperature) of the elastic moduli and seismic wave velocities, which are the main subject of this work.

Seismic tomography gives insight into the three-dimensional structure and dynamics of the Earth, but its quantitative interpretation requires knowledge of temperature variations of seismic wave

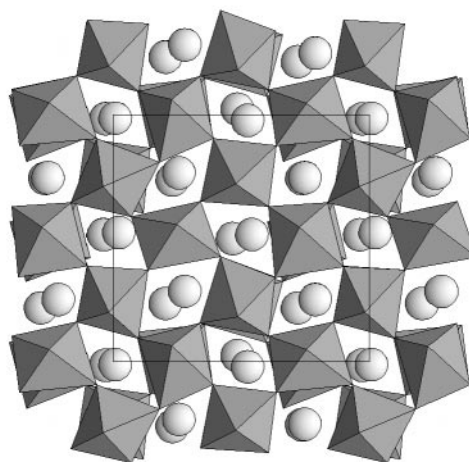


Figure 1 'Snapshot' of the crystal structure of MgSiO_3 perovskite from *ab initio* molecular dynamics simulations at 88 GPa and 3,500 K. The equilibrium structure at these conditions is orthorhombic. Spheres, Mg atoms; polyhedra SiO_6 octahedra. The supercell used in the calculations ($2a \times 2b \times 1c$ *Pbnm*-supercell, contours shown) contains 80 atoms. Simulations were also performed at 38 GPa and 1,500 K, 38 GPa and 2,500 K, 38 GPa and 3,500 K, 88 GPa and 1,500 K, and 136 GPa and 4,500 K, and at all conditions the optimized cell was orthorhombic. The simulated pressures of 38 GPa and 88 GPa correspond to depths of approximately 1,000 km and 2,000 km, respectively. Our simulations are based on density functional theory^{23,24} within the generalized gradient approximation²⁵ and pseudopotential plane wave scheme. The pseudopotentials used are non-local, ultrasoft for O and norm-conserving with partial core corrections for Mg and Si. Plane-wave cut-off of 500 eV, which gives excellent convergence of all properties with respect to the basis set size, was used. Supercell periodicity was imposed on the wavefunction, which proved to give sufficiently accurate results with our large supercell. At each pressure/temperature condition, the simulations were run in the constant-*NVT* ensemble²⁶ for at least 0.79 ps after equilibration with 1 fs timestep. This was sufficient to produce accurate statistical averages of the stress tensor components. Cell dimensions were carefully optimized by making the stress tensor hydrostatic.

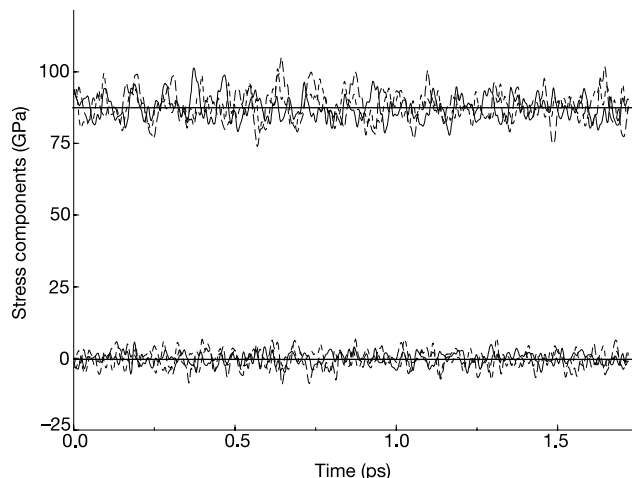


Figure 2 Fluctuations of the stress tensor components for the structure shown in Fig. 1. Data for the optimized cell at 88 GPa and 3,500 K show that the stress is hydrostatic. As the generalized gradient approximation often results in a shifted pressure scale¹⁴, we corrected the calculated pressures and diagonal stress components by -12.08 GPa. This correction alters only the pressure scale, but not the properties. More information on the pressure correction and performance of the generalized gradient approximation can be found in ref. 14. Applying strains (we used one triclinic and three axial strains with

velocities. Seismology gives $R_T = (\partial \ln v_s / \partial \ln v_p)_T$, the ratio of variation of the shear (v_s) and compressional (v_p) velocities due to pressure alone, as 0.7 (ref. 1). A similarly defined parameter, $R_p = (\partial \ln v_s / \partial \ln v_p)_p$ measuring the same ratio, but due to temperature effects alone, is much larger: it increases from 1.7 to 2.6 between the depths of 1,000 km and 2,000 km according to seismic tomography data¹⁵. The large difference between R_T and R_p has been a puzzle for geophysicists over past 15 years; Anderson¹ suggested that the difference is entirely due to intrinsic anharmonicity, the simulation of which needs molecular dynamics or Monte Carlo methods, rather than the quasiharmonic approximation.

Our *ab initio* molecular dynamics simulations (Figs 1 and 2) were performed with the VASP code³ run on 128 nodes of the CRAY T3E supercomputer at Manchester Computer Centre. The calculated equilibrium lattice parameters, isothermal and adiabatic elastic

magnitudes $+0.02$ and -0.02) and calculating the time-averages of the induced stresses, we calculated the isothermal elastic constants C_{ij} from nonlinear stress-strain relations; different elastic constants are obtained from different stresses. Adiabatic elastic constants, used in calculations of seismic velocities, were calculated from the isothermal constants²⁷ using the thermal pressure tensor calculated directly by *ab initio* molecular dynamics.

constants, and thermal expansion coefficients are presented in Table 1. One of the calculated points, 38 GPa and 3,500 K, is close to the melting point of MgSiO_3 perovskite¹⁶ (results of this study indicated that melting would occur at $\sim 3,800$ K under a pressure of 38 GPa; however, there is still controversy about the melting curve of MgSiO_3 perovskite). Although we observed very large displacements of atoms, all atoms vibrated about their ideal crystallographic locations and did not diffuse, and shear elastic constants were large and positive. It is remarkable that even at these conditions orthorhombic MgSiO_3 perovskite does not transform to a cubic or tetragonal phase. Anomalous elastic constants obtained at these critical conditions were not used in the further analysis.

We calculate $R_T = 0.61$ – 0.81 (our static simulations¹⁴ give 0.59–0.74), in keeping with geophysical observations. The value of R_p increases from 1.5 at a depth of 1,000 km to 1.9 at 2,000 km. These

Table 1 Properties of MgSiO_3 perovskite at high pressures and temperatures

Property	Conditions				
	38 GPa, 1,500 K	38 GPa, 2,500 K	38 GPa, 3,500 K	88 GPa, 1,500 K	88 GPa, 3,500 K
V	147.63	150.83	154.08	132.55	136.48
a	4.628	4.675	4.723	4.435	4.499
b	4.792	4.814	4.836	4.655	4.674
c	6.658	6.703	6.747	6.420	6.492
$\sigma_{1,2,3}$	37.9; 38.0; 37.9	38.0; 37.6; 37.3	37.0; 39.0; 38.0	88.1; 87.7; 88.4	87.8; 87.9; 88.5
P	37.9	37.6	38.0	88.1	88.1
C_{11}/T	601/616	553/579	525/564	813/826	749/783
C_{22}/T	697/711	629/654	545/578	978/991	847/878
C_{33}/T	649/663	591/615	545/580	933/945	821/850
C_{12}/T	266/280	233/258	222/258	464/477	398/431
C_{13}/T	235/249	218/243	218/256	348/362	324/356
C_{23}/T	251/264	240/265	251/286	383/396	364/395
C_{44}	262	232	202	336	270
C_{55}	219	210	180	266	234
C_{66}	199	178	147	264	195
$K_{T/S}$	382.7/396.8	349.9/375.8	332.7/369.5	565.7/579.4	508.6/542.5
G	214.3	194.9	166.2	269.7	226.9
γ	1.30	1.37	1.40	1.21	1.26
α	1.91	2.15	2.26	1.34	1.51

Here V , a , b , c are unit cell volume and lattice parameters, in $\text{\AA}^3$ and $\text{\AA}$, respectively; $\sigma_{1,2,3}$ are stress tensor components, showing that non-hydrostatic deviations from the average pressure P are very small; C_{ij} are elastic constants (T , isothermal; S , adiabatic) in GPa (errors of a few per cent); K and G are Voigt-Reuss-Hill bulk and shear moduli in GPa (errors within 2%); γ is the Grüneisen parameter and α thermal expansion in 10^{-5} K^{-1} (errors within 10%). Important derivatives: $(\partial \ln v_s / \partial T)_p = -3.67 \times 10^{-5} \text{ K}^{-1}$ (38 GPa), $-3.78 \times 10^{-5} \text{ K}^{-1}$ (88 GPa), $(\partial \ln v_p / \partial T)_p = -2.49 \times 10^{-5} \text{ K}^{-1}$ (38 GPa), $-1.98 \times 10^{-5} \text{ K}^{-1}$ (88 GPa), $(\partial \ln v_s / \partial T)_p = -1.65 \times 10^{-5} \text{ K}^{-1}$ (38 GPa), $-0.92 \times 10^{-5} \text{ K}^{-1}$ (88 GPa), $(\partial \ln v_s / \partial \ln v_p)_T = 0.61$ (1,500 K), 0.81 (3,500 K), $(\partial \ln v_s / \partial \ln v_p)_p = 1.48$ (38 GPa), 1.91 (88 GPa). All errors are within 20%.

values, obtained taking full account of anharmonic effects, are still lower than the observed values. Karki *et al.*¹⁷, using the quasiharmonic approximation, found R_p of MgO to increase from 1.4 to 1.9 from the top to the bottom of the lower mantle. This agrees well with our results for MgSiO₃; the remaining deficit of R_p can only be explained by either significant anelasticity⁴ or large-scale compositional heterogeneity of the lower mantle.

Seismic tomography is a quickly developing technique; the most recent tomography maps have similar qualitative features (locations of cold slabs and hot plumes), and give the absolute velocity perturbations within ~25% uncertainty. For interpreting seismic tomography in terms of temperature, it is most convenient to use the bulk sound velocity v_ϕ , whose seismological determination is unaffected by anelastic effects. The bulk velocity maps can be accurately determined by the joint inversion of the shear and compressional velocities. Perhaps the most reliable global seismic tomography maps currently available are those of Masters *et al.*⁵. Older higher-resolution maps of Kennett *et al.*¹⁸ realistically give much narrower cold (high-velocity) zones, but strongly underestimate the amplitudes of the velocity variations.

We find the temperature derivative of the bulk velocity, $\phi = (\partial \ln v_\phi / \partial T)_p$, to vary from $-1.7 \times 10^{-5} \text{ K}^{-1}$ at 1,000 km to $-0.9 \times 10^{-5} \text{ K}^{-1}$ (errors $\pm 20\%$) at 2,000 km depth. Temperature variations can be obtained from the relative velocity perturbations: $\delta T = (\delta v_\phi / v_\phi) / \phi$. Combined with bulk sound velocity maps of Masters *et al.*⁵ (with maximum velocity contrast $\Delta v_\phi / v_\phi = 1.4\%$ at 1,225 km and 2,195 km), this gives the temperature contrast between the hot plumes and cold slabs increasing from 800 K to 1,500 K between 1,000 and 2,000 km depth. Linear extrapolation gives ~550 K at the top of the lower mantle, and 2,150 K at the bottom. The root-mean-square temperature variations are ~150–250 K across the lower mantle. Our temperature contrasts are ~2–4 times smaller than estimates¹⁹ obtained using extremely uncertain extrapolations of $s = (\partial \ln v_s / \partial T)_p$ and which would suggest partial melting of the lower mantle. Our results can support partial melting only in the lowermost part of the lower mantle. Castle *et al.*²⁰, using their shear velocity maps for the bottom of the lower mantle (maximum $\Delta v_s / v_s = 7.4\%$) and a reasonable guess for s , obtained an estimate of the temperature contrast at the core–mantle boundary similar to ours, ~1,600 K. Using the v_ϕ -maps of Kennett *et al.*¹⁸ (maximum $\Delta v_s / v_s = 1.0\%$ at 1,225 km and 2,195 km), we get much smaller temperature contrasts, rising from ~500 K to ~1,000 K between the depths of 1,000 km and 2,000 km.

It is becoming increasingly clear that temperature variations are the main factor determining the perturbations of seismic wave velocities in the lower mantle. Anelasticity can be important for the shear velocities, but is negligible for the bulk velocities. Compositional heterogeneity is significant only near the core–mantle boundary⁵; it can occur due to the chemical reaction²¹ between core and mantle, driving Fe from the silicate mantle into the metallic core. Hot plumes, rising from the core–mantle boundary, would then be depleted in Fe. When substituting for Mg, Fe dramatically decreases shear moduli of many silicates and oxides, while the effect on bulk moduli is very small. Hence, even minor regional variations of the Fe content can have a major effect on the shear velocities, being less important for the bulk velocity. For example, very hot, slightly Fe-depleted mantle rocks can have the same shear velocities as very cold, slightly Fe-enriched mantle rocks. Consistent with the depletion of plumes in Fe, we find the temperature contrasts at 2,000 km determined for the shear velocities^{5,18} (1,000 K from maps⁵) to be smaller than those determined from the corresponding bulk velocities (1,500 K from maps⁵). This effect can even produce an anticorrelation between the bulk and shear velocities, but not its observed⁵ pattern: it remains a mystery why it is the bulk (not shear) velocity anomalies that undergo a reversal near the core–mantle boundary (that is, low-velocity zones at the core–mantle boundary underlay high-velocity anomalies of the rest of the

lower mantle). In any case, the temperature contrasts obtained from the bulk velocities must be the most reliable, at least outside the anomalous core–mantle boundary region, in which compositional effects seem to be significant for all types of seismic waves. The temperature contrasts that we have found could be used as important constraints in numerical modelling of mantle convection.

Kesson *et al.*²² estimated that lithospheric slabs should be at least ~650 K colder than surrounding mantle if they are to sink to the core–mantle boundary, and at least 250 K colder to reach the depth of 1,100 km. Our maximum cold temperature anomalies (roughly half of the total temperature contrasts) are similar to these estimates, and suggest that some lithospheric slabs might stop sinking before reaching the core–mantle boundary. Neutrally buoyant slabs would be dissolved by the convecting mantle; some tomographic maps show most slabs disappearing in the middle of the lower mantle¹⁸, while others⁵ show that most slabs do reach the core–mantle boundary.

The next step is to estimate the extent of chemical heterogeneity and anelasticity in the lower mantle, and construct a three-dimensional mineralogical model of the Earth's mantle. With improved seismic tomography models and mineral physics data, this could be achieved in the foreseeable future. We believe that *ab initio* molecular dynamics simulations will be important in solving this and many other geologically important problems. □

Received 21 September 2000; accepted 30 April 2001.

- Anderson, D. L. *Theory of the Earth* (Blackwell Scientific Publications, Boston, 1989).
- Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
- Kresse, G. & Furthmüller, J. Efficiency of *ab initio* total-energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mater. Sci.* **6**, 15–50 (1996).
- Karato, S.-i. Importance of anelasticity in the interpretation of seismic tomography. *Geophys. Res. Lett.* **20**, 1623–1626 (1993).
- Masters, G., Laske, G., Bolton, H. & Dziewonski, A. In *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale* (eds Karato, S.-i. *et al.*) 63–87 (AGU Geophys. Monogr. 113, American Geophysical Union, Washington DC, 2000).
- Ancilotto, F., Chiarotti, G. L., Scandolo, S. & Tosatti, E. Dissociation of methane into hydrocarbons at extreme (planetary) pressure and temperature. *Science* **275**, 1288–1290 (1997).
- Cavazzoni, C. *et al.* Superionic and metallic states of water and ammonia at giant planet conditions. *Science* **283**, 44–46 (1999).
- Alfé, D., Gillan, M. J. & Price, G. D. The melting curve of iron at the pressures of the Earth's core from *ab initio* calculations. *Nature* **401**, 462–464 (1999).
- Sugino, O. & Car, R. *Ab initio* molecular dynamics study of first-order phase transitions: melting of silicon. *Phys. Rev. Lett.* **74**, 1823–1826 (1995).
- Buda, F., Car, R. & Parrinello, M. Thermal expansion of c-Si via *ab initio* molecular dynamics. *Phys. Rev. B* **41**, 1680–1683 (1990).
- da Silva, C., Wentzcovitch, R. M., Patel, A., Price, G. D. & Karato, S.-i. The composition and geotherm of the lower mantle: constraints from the elasticity of silicate perovskite. *Phys. Earth Planet. Inter.* **118**, 103–109 (2000).
- O'Keeffe, M. & Bovin, J. O. Solid electrolyte behavior of NaMgF₃: geophysical implications. *Science* **206**, 599–600 (1979).
- Matsui, M. & Price, G. D. Simulation of the pre-melting behaviour of MgSiO₃ perovskite at high pressures and temperatures. *Nature* **351**, 735–737 (1991).
- Oganov, A. R., Brodholt, J. P. & Price, G. D. *Ab initio* elasticity and thermal equation of state of MgSiO₃ perovskite. *Earth Planet. Sci. Lett.* **184**, 555–560 (2001).
- Robertson, G. S. & Woodhouse, J. H. Constraints on lower mantle properties from seismology and mineral physics. *Earth Planet. Sci. Lett.* **143**, 197–205 (1996).
- Shen, G. & Lazor, P. Measurement of melting temperatures of some minerals under lower-mantle pressure. *J. Geophys. Res.* **100**, 17699–17713 (1995).
- Karki, B. B., Wentzcovitch, R. M., de Gironcoli, S. & Baroni, S. First-principles determination of elastic anisotropy and wave velocities of MgO at lower mantle conditions. *Science* **286**, 1705–1707 (1999).
- Kennett, B. L. N., Widiyantoro, S. & van der Hilst, R. D. Joint seismic tomography for bulk sound and shear wave speed in the Earth's mantle. *J. Geophys. Res.* **103**, 12469–12493 (1998).
- Yuen, D. A., Cadek, O., Chopelas, A. & Matyska, C. Geophysical inferences of thermal-chemical structures in the lower mantle. *Geophys. Res. Lett.* **20**, 899–902 (1993).
- Castle, J. C., Creager, K. C., Winchester, J. P. & van der Hilst, R. D. Shear wave speeds at the base of the mantle. *J. Geophys. Res.* **105**, 21543–21557 (2000).
- Knittle, E. & Jeanloz, R. Earth's core–mantle boundary: results of experiments at high pressures and temperatures. *Science* **251**, 1438–1443 (1991).
- Kesson, S. E., Fitz Gerald, J. D. & Shelley, J. M. Mineralogy and dynamics of a pyrolite mantle. *Nature* **393**, 252–255 (1998).
- Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
- Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
- Wang, Y. & Perdew, J. P. Correlation hole of the spin-polarized electron gas, with exact small-vector and high-density scaling. *Phys. Rev. B* **44**, 13298–13307 (1991).

26. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
 27. Wallace, D. C. *Thermodynamics of Crystals* (Dover Publications, New York, 1998).

Acknowledgements

We acknowledge receipt of a Russian President Scholarship for Education Abroad, a UCL Graduate School Research Scholarship, and a UK Overseas Research Scholarship (A.R.O.) and a Royal Society University Research Fellowship (J.P.B.). We thank NERC for access to the supercomputer facilities.

Correspondence and requests for materials should be addressed to A.R.O. (e-mail: a.oganov@ucl.ac.uk).

Major fungal lineages are derived from lichen symbiotic ancestors

François Lutzoni¹*, Mark Pagel²‡ & Valérie Reeb¹*†§

^{*} Department of Botany, The Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, Illinois 60605, USA

[‡] School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK

[§] Department of Biological Sciences, University of Illinois at Chicago (M/C 066), Chicago, Illinois 60607, USA

About one-fifth of all known extant fungal species form obligate symbiotic associations with green algae, cyanobacteria or with both photobionts. These symbioses, known as lichens, are one way for fungi to meet their requirement for carbohydrates^{1,2}. Lichens are widely believed to have arisen independently on several occasions, accounting for the high diversity and mixed occurrence of lichenized and non-lichenized (42 and 58%, respectively) fungal species within the Ascomycota^{3,4}. Depending on the taxonomic classification chosen^{2,5,6}, 15–18 orders of the Ascomycota include lichen-forming taxa, and 8–11 of these orders (representing about 60% of the Ascomycota species) contain both lichenized and non-lichenized species. Here we report a phylogenetic comparative analysis of the Ascomycota, a phylum that includes greater than 98% of known lichenized fungal species⁵. Using a Bayesian phylogenetic tree sampling methodology^{7,8} combined with a statistical model of trait evolution⁹, we take into account uncertainty about the phylogenetic tree and ancestral state reconstructions. Our results show that lichens evolved earlier than believed, and that gains of lichenization have been infrequent during Ascomycota evolution, but have been followed by multiple independent losses of the lichen symbiosis. As a consequence, major Ascomycota lineages of exclusively non-lichen-forming species are derived from lichen-forming ancestors. These species include taxa with important benefits and detriments to humans, such as *Penicillium* and *Aspergillus*^{10–12}.

To investigate the evolution of the lichen symbiosis it is necessary to account for the phylogenetic relationships within the Ascomycota, and to infer the rates and likely pattern of gains and losses of the symbiotic state. We reduced the high level of uncertainty associated with small subunit nuclear ribosomal RNA gene (SSU nuclear rDNA) phylogenies of the Ascomycota^{13–15}, by obtaining sequences from the small and large subunit (LSU) of the nuclear rRNA genes for 52 species of the Ascomycota. Our sample includes representatives from 24 of 46 orders², representing ≈ 75% of the Ascomycota species diversity.

Phylogenetic comparative investigations typically rely on a single phylogenetic tree and reconstruct ancestral states on the basis of the method of parsimony. However, phylogenetic trees are rarely known without error and different tree topologies can give different estimates of ancestral states. In addition, ancestral states reconstructed by parsimony do not account for the statistical uncertainty of ancestral inferences. Both of these problems are acute when reconstructing the evolution of the lichen symbiosis. All previous broad phylogenetic studies of the Ascomycota had low bootstrap support and unstable relationships in critical portions of the trees^{4,12–17}. Furthermore, parsimony methods may perform poorly when rates of character evolution are high^{18,19}.

We used a Bayesian statistical procedure based on Markov chain Monte Carlo (MCMC) sampling methods⁷ to account for phylogenetic uncertainty. This sampling procedure allows us to draw a random sample from the universe of possible phylogenetic trees. The frequency distribution of the sample estimates the posterior probability distribution of trees (see Methods). From the distribution of sampled trees we calculated the posterior probability of ancestral nodes and focused our data interpretation on those nodes with the highest statistical certainty.

We used a statistical model of trait evolution^{9,20,21} to estimate on each tree the evolutionary rate of gains and losses of lichenization, and the most probable ancestral states (lichen-forming/non-lichen-forming) at specified nodes. Rates of evolution between states are calculated over all possible states at each node of a given tree and are therefore independent of any particular reconstruction of the ancestral states. The model of trait evolution takes into account the lengths of the branches of the phylogenetic tree, does not constrain the rates of gains and losses *a priori*, and expresses the statistical uncertainty associated with estimates of ancestral states at each node. To derive an overall estimate of the rates of evolution or the probability of an ancestral state, the estimates from each tree are averaged (see Methods). Our inferences about the nature of the evolutionary processes underlying lichen evolution thereby take account of uncertainty inherent to the phylogenetic hypothesis, and are not conditional on any particular tree.

We sampled 19,900 phylogenetic trees using the MCMC procedure⁷, and estimated by maximum likelihood the rates of gains and losses of lichenization on each (Fig. 1). Larger rate values correspond to a higher expected number of transformations (losses or gains), and the loss/gain ratio (dots in Fig. 1) directly estimates the ratio of expected losses to expected gains of lichenization during evolution.

In contrast with previous work on the evolution of the lichen symbiosis⁴, our results show that rates of loss of lichenization exceed rates of gain in the Ascomycota. In 18,029 of the 19,900 sampled trees (90.6%) the estimated rate of loss exceeds the rate of gain (that is, the loss/gain ratio is greater than one, and therefore is above the

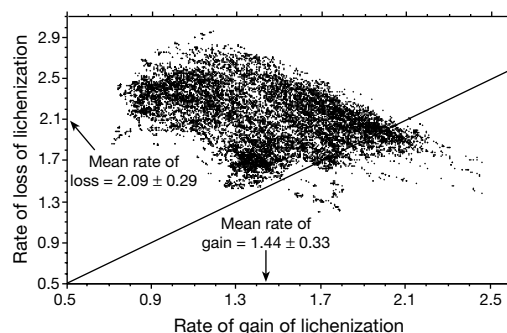


Figure 1 The rate of loss of lichenization exceeds the rate of gain of lichenization, independently of tree topology. Data are for 19,900 MCMC trees. The 1:1 relationship is indicated by the solid line.

† Present address: Department of Biology, Duke University, Durham, North Carolina 27708, USA.

diagonal line in Fig. 1). The average loss/gain ratio is 1.56 ± 0.53 and is skewed towards higher ratios (range = 0.56–3.24). Across trees the highest rates of loss are associated with the lowest rates of gain (Fig. 1; $r = -0.40$). The ratio of losses to gains is, however, independent of the phylogenetic tree topology (correlation between ratio of rate of losses to gains and log likelihood of trees = -0.024).

These results indicate that there have been at least 1.5 times as many losses of the lichen symbiotic state as gains during the evolution of the Ascomycota. The majority rule consensus phylogenetic tree (Fig. 2, right) shows the most probable occurrence of gains and losses of lichenization and their phylogenetic context. We show this tree not to propose a particular topology, but principally

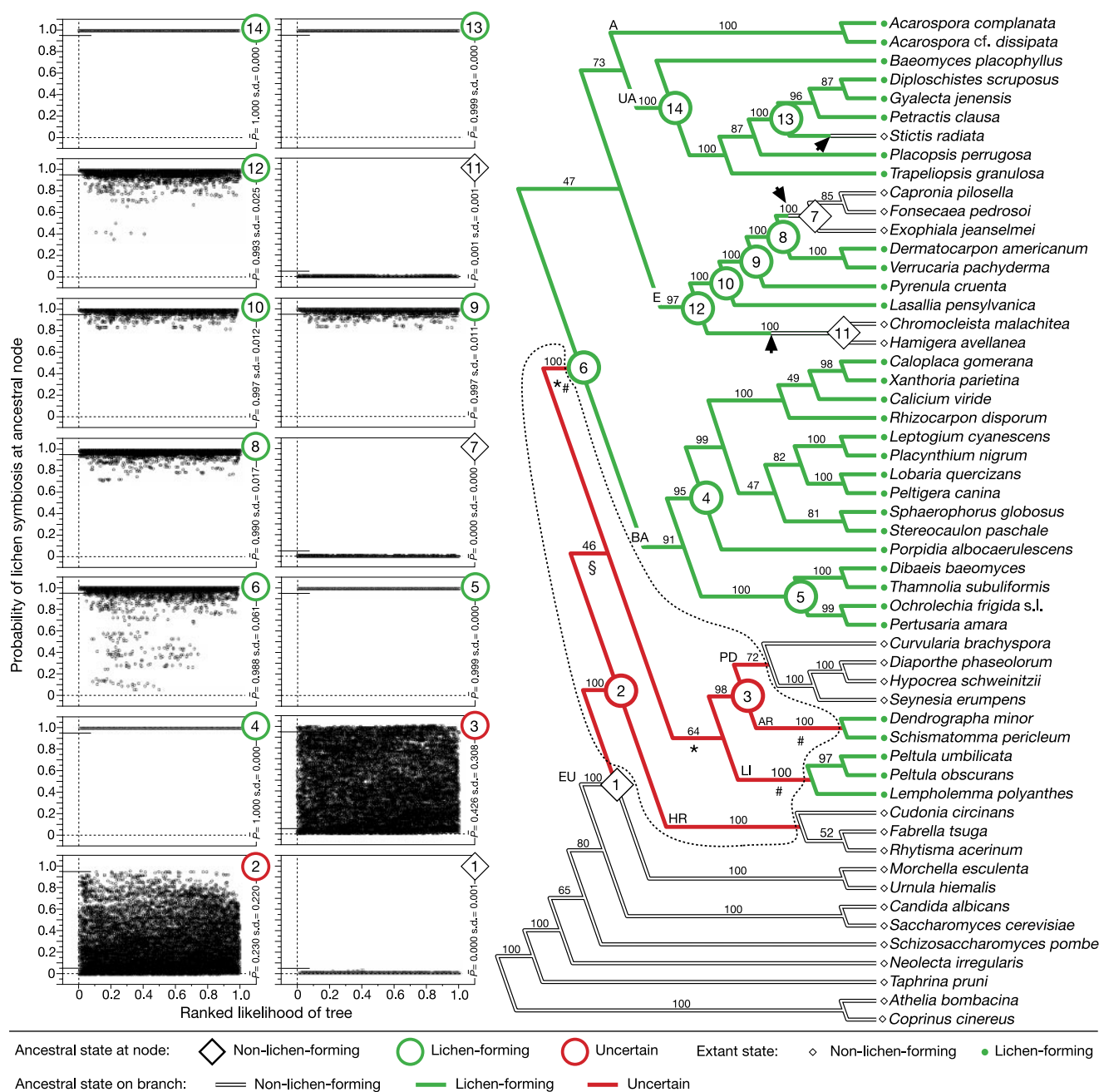


Figure 2 Bayesian posterior probabilities for reconstructed evolution of the lichen symbiosis and for each node of the Ascomycota phylogeny. Numbers (1–14) connect nodes in the tree with their respective graphs. Left, reconstructed probability (Discrete^{3,20}) that ancestral state was lichen-forming at specified node on 19,900 trees generated by MCMC sampling (BAMBE⁷). The average probability and standard deviation as calculated across all trees is provided on the y-axis to the right of each graph. Right, Ascomycota majority rule consensus of 19,900 MCMC-sampled trees on the basis of SSU and LSU nuclear rDNA sequences. The number above each branch corresponds to the posterior probability (%) of the node to which it points. The region of the tree for which the ancestral states of branches could not be extrapolated, because of uncertainty associated with specific nodes, is indicated by a dotted line. Arrows indicate losses of lichenization. The

non-lichenized Archiascomycetes (subphyllum Taphrinomycotina) are represented by the paraphyletic *Taphrina*, *Schizosaccharomyces* and *Neolecta*. The non-lichenized Hemiascomycetes (Saccharomycotina), are represented by *Saccharomyces* and *Candida*. The operculate discomycetes (Pezizomycotina), represented by *Uromyces* and *Morchella*, form the first (basal) divergence within the highly stable Euscomycetes (EU). The Acarosporaceae (A), 'unitunicate ascohymeniales' (UA), 'bitunicate ascohymeniales' (BA) and 'Eurotiomycotina' (E) consistently formed a monophyletic group—the 'Lecanoromycotina' (node 6). The Helotiales–Rhytismatales (HR), Lichinales (LI), Arthoniales (AR) and Pyrenomycetes–Dothideales (PD) appear as a paraphyletic assemblage with poor support.

to provide a means for discussing central events of evolution in the Ascomycota. The numbers above each internal branch of the consensus tree (Fig. 2) correspond to the Bayesian posterior probability of the node to which a branch points. Nodes with values of 100 define a collection of species all of which—and only those of which—appeared in every one of the trees sampled from the Markov chain. Other nodes were less certain, and emphasize the need for a statistical approach.

We restricted our reconstructions⁹ of the probable ancestral states (lichen-forming/non-lichen-forming) to fourteen nodes, each of which gained a posterior probability of $\geq 95\%$ in our MCMC sample (numbers 1–14; Fig. 2, right). These nodes delineate groups of lichen-forming and non-lichen-forming species in such a way as to make it possible to put reasonable upper and lower limits on the number of independent gains and losses of lichenization. The left panel of Fig. 2 plots the probability of the ancestral state for each of these labelled nodes as reconstructed across the 19,900 sampled trees.

The ancestor at 9 of the 14 nodes is lichenized (green circles, corresponding to > 0.98 posterior probability of reconstructed state; Fig. 2), 3 nodes are non-lichenized (black diamonds, corresponding to < 0.01 posterior probability), and the symbiotic status for two ancestral nodes is uncertain (red circles, corresponding to a 0.23 and 0.43 probability of being lichen-forming). Thus, green areas of the tree are regions of lichenization, red areas are regions in which the ancestral state is uncertain, and the remaining (uncoloured) branches correspond to non-lichenized regions.

The pattern of ancestral states we find can be used to infer that a minimum of one and a maximum of three gains of lichenization occurred during the evolution of the Ascomycota, with perhaps one or two being the most likely. If lichen formation originated immediately after node 2 (branch labelled with \$), then one gain of lichenization is implied for the Ascomycota. Two origins of lichen symbiosis are implied (branches labelled with *) if lichenization evolved independently at node 6 and again at the base of the clade that includes the Lichinales (LI), Arthoniales (AR) and Pyrenomyces–Dothideales (PD). Three independent gains are implied if the closely related AR and LI groups independently evolved lichenization (branches labelled with #).

By comparison, a minimum of three and possibly four losses of lichenization have occurred in the Ascomycota. Nodes 8, 12 and 13 each have very high posterior probabilities on the tree (Fig. 2, right), each is reconstructed to have a high posterior probability of being lichenized (Fig. 2, left), and each is followed by a loss of lichenization on the tree. The non-lichenized members of the Ostropales (such as *Stictis radiata*, derived from node 13), the Chaetothyriales (all species derived from node 7), and the Eurotiales (node 11, which now includes the Onygenales^{10,12}) are therefore all secondarily non-lichenized, being independently derived from lichenized ancestors. The Eurotiales and Onygenales together encompass about 320 species, none of which are lichenized²; the Chaetothyriales (≈ 75 species) are not known to have lichenized species; and the Ostropales (including Graphidales) comprise about 1,850 species, 10% of which are non-lichenized (≈ 180 species). A fourth loss of lichenization is implied at the base of the PD group if lichen formation indeed originated only once or twice within the Ascomycota (that is, at the branches labelled with \$ or with *; Fig. 2).

These results reshape in three related and important ways our understanding of the evolution of the lichen symbiosis in the Ascomycota. First, we find that lichens arose much earlier (node 6; Fig. 2) than previously thought. This suggests that the lichen symbiosis has been a relatively long standing relationship, and consequently, has had a larger role in the evolution of the Ascomycota than previously believed⁴. Second, several major lineages of strictly non-lichenized species (Chaetothyriales and Eurotiales) unexpectedly turn out to be derived from lichen-forming ancestors. If Fig. 2 is extrapolated to all Ascomycota species, then a minimum

of about 4% of known extant non-lichenized Ascomycota species are secondarily derived from lichen symbiotic ancestors. The actual percentage could be much higher, as the non-lichenized species within the Arthoniales (AR) and Pyrenomyces–Dothideales (PD) also may have resulted from an ancestral loss of lichenization.

A third implication of our results is to emphasize the distinction between secondarily derived and primary non-lichenized Ascomycota. Candidiasis, for example, is caused by the primary non-lichenized fungus *Candida albicans*. A large number of Ascomycota species of relevance to humans, however, are shown here to be those that have secondarily lost the ability to form a lichen symbiosis. The Eurotiales (node 11), for example, include many of the most beneficial and detrimental fungi to humans because of their production of antibiotics and mycotoxins¹¹. *Penicillium* is a member of the secondarily derived non-lichenized Eurotiales. Aflatoxins produced by *Aspergillus flavus* and *Aspergillus parasiticus* (also members of the Eurotiales) on various nuts and grains are among the most potent carcinogenic compounds known¹¹. Other attributes of members of the Eurotiales include being used in the ripening of blue cheeses, being pathogens of citrus fruits, and being infective agents of animal and human diseases such as aspergilloses, caused by the secondarily derived non-lichenized *Aspergillus*¹¹. The Herpotrichiellaceae (Chaetothyriales) is another example of a major fungal lineage derived from a lichenized ancestor that gave rise to opportunistic pathogens to humans. *Fonsecaea pedrosoi* (node 7; Fig. 2, right) is one of the causative agents of chromoblastomycosis, and certain species of *Exophiala* are pathogens of fish, including salmon¹¹. Better insight into animal and human diseases caused by fungi may be gained by calling attention to the distinction between secondarily derived and primary non-lichen-forming taxa.

Secondarily derived non-lichen-forming fungi may be more likely to synthesize compounds that have medicinal, pharmacological, carcinogenic and food production properties because the original transition to lichenization may have shifted the selective pressure on biosynthetic pathways towards an accelerated origin of new secondary compounds. It is well known that lichen-forming Ascomycota are prolific in the production of unique secondary compounds (especially depsides and depsidones), some of which have antibiotic or anti-tumour properties²². It is expected that when lichen formation is lost, some of the genes involved in these biosynthetic pathways may be diverted to new functions in the secondarily derived non-lichen-forming organism. Phylogenies can help to identify additional species with possible benefits to humans.

Lichenicolous fungi (fungi dwelling on or in lichens as parasites, commensals or saprobes²) may provide one explanation for the high rate of loss of the lichen symbiotic habit reported here (Figs 1 and 2). Morphological evidence supports the hypothesis that many of these lichenicolous fungi are derived from lichen ancestors²³. By shedding their lichenized habit and colonizing lichens, these non-lichenized fungi continue to obtain directly, or indirectly from the lichenized fungus, carbohydrates generated by the alga/cyanobacterium without having to find a specific free living photobiont with which to form a lichen thallus *de novo* each generation²³. This transformation from a lichenized to a non-lichenized lichenicolous state may act as a fungal 'half-way house' that could facilitate further transitions to different substrates. The high rate of loss of lichenization that we have found corroborates a study that mutualism is not necessarily at the end of a unidirectional evolution from parasitism to mutualism²⁴.

A phylogenetic study of the basidiomycete *Omphalina*, which includes a mixture of closely related lichenized and non-lichenized species²⁵, and where a single unequivocal gain of lichenization was detected²⁶, supports the low rate of gains of lichenization reported here. We have shown elsewhere²⁷ that this specific transition to mutualism leads to accelerated rates of evolution at the molecular level. Slow growth in axenic culture of these lichen-forming mushrooms, the unusual high frequency of uninucleated individuals, and

the highly variable number of spores/basidium compared with their non-lichenized sister species, are further evidence that a high level of stress on the fungus is associated with a transition to the lichen symbiosis²⁸. This may suggest that many attempts to form stable lichen symbioses occur in nature, but only rarely does a specific fungal lineage have all the requirements to survive the costs associated with a transition to this symbiotic state. □

Methods

DNA sequencing, taxon and character sampling

Total DNA was isolated, and the SSU and LSU nuclear rDNA were amplified, sequenced and aligned as described in ref. 16. Regions of the alignments where the placements of gaps were ambiguous were removed from the MCMC phylogenetic tree sampling analyses. Taxa were selected to represent all main ascomycete orders²⁹ known to include lichenized species (13 out of 15 orders) and nearly all main orders of Ascomycota known to include only non-lichenized species (11 out of 31 orders). At least 16 of the unsampled non-lichenized orders almost certainly fall entirely within existing non-lichenized clades (Fig. 2) and their sampling will not affect the results presented here. This is because the reconstruction of ancestral states is not weighted by the number of descendant taxa that have a particular state; rather, the reconstructed state depends on the relative frequencies of the states in the descendants and their phylogenetic distribution. If all of a group of unsampled taxa share a most recent common ancestor and the same state with a species already included in our study, our reconstructions are unchanged. Basidiomycota (*Athelia bombacina* and *Coprinus cinereus*) sequences were included as outgroups. (The voucher/GenBank information is available as Supplementary Information.) We generated a total of 20 SSU and 24 LSU nuclear rDNA sequences in this study. All these sequences were deposited in GenBank under accession numbers AF356653–AF356696. Ten SSU and 20 LSU sequences were from ref. 16, and the remaining 24 SSU and 10 LSU sequences were from GenBank.

MCMC phylogenetic tree sampling

We used MCMC methods⁷ within a Bayesian framework to estimate the posterior probability of phylogenetic trees. The MCMC procedure ensures that trees are sampled in proportion to their probability of occurrence under the model of gene-sequence evolution. We generated 200,000 phylogenetic trees using the MCMC procedure, sampling every tenth one to assure that successive samples were independent^{27,29}. We then removed the first 100 trees in the sample to avoid including any trees that might have been sampled before convergence of the Markov chain. We used the general time-reversible model of gene-sequence evolution combined with gamma rate heterogeneity to estimate the likelihood of each tree³⁰. Information on the state of each species (lichen-forming/non-lichen-forming) was excluded from the MCMC sampling procedure to ensure that the distribution of tree topologies was not influenced by this trait. A series of runs using the BAMBE⁷ 'global' and 'local' options was conducted to ensure that the Markov chain converged to the same region in the universe of trees.

Reconstruction of gains and losses, and ancestral states

We used a continuous time Markov model of trait evolution, as implemented in the computer program Discrete (available from M.P.), allowing independent gains and losses in each branch of the phylogenetic tree²⁹. Parameters specifying rates of gain (q_{01}) and loss (q_{10}) of lichenization were calculated separately for each tree sampled in the MCMC procedure, following procedures in ref. 9. Because trees are represented in the sample in proportion to their likelihood, investigating the rates over all trees automatically weights our results by the likelihood of a particular tree type. A detailed description of the analyses performed for this study will be published as a book chapter (M.P. and F.L., manuscript in preparation).

Received 21 December 2000; accepted 30 April 2001.

- Hawksworth, D. L. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycol. Res.* **95**, 641–655 (1991).
- Hawksworth, D. L., Kirk, P. M., Sutton, B. C. & Pegler, D. N. *Dictionary of the Fungi* (CAB, Wallingford, 1995).
- Barinaga, M. Origins of lichen fungi explored. *Science* **268**, 1437 (1995).
- Gargas, A., DePriest, P. T., Grube, M. & Tehler, A. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. *Science* **268**, 1492–1495 (1995).
- Honegger, R. in *Lichen Biology* (ed. Nash, T. H.) 24–36 (Cambridge Univ. Press, Cambridge, 1996).
- Hafellner, J. in *Handbook of Lichenology* Vol. 3 (ed. Galun, M.) 41–52 (CRC, Boca Raton, 1988).
- Larget, B. & Simon, D. L. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16**, 750–759 (1999).
- Huelsenbeck, J. P., Rannala, B. & Masly, J. P. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* **288**, 2349–2350 (2000).
- Pagel, M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48**, 612–622 (1999).
- Berbee, M. L., Yoshimura, A., Sugiyama, J. & Taylor, J. W. Is *Penicillium* monophyletic? An evaluation of phylogeny in the family Trichocomaceae from 18S, 5.8S, and ITS ribosomal DNA sequence data. *Mycologia* **87**, 210–222 (1995).
- Alexopoulos, C. J., Mims, C. W. & Blackwell, M. *Introductory Mycology* (John Wiley & Sons, New York, 1996).
- Suh, S.-O. & Blackwell, M. Molecular phylogeny of the cleistothecial fungi placed in Cephalothecaceae and Pseudoeurotiaceae. *Mycologia* **91**, 836–848 (1999).

- Berbee, M. L. & Taylor, J. W. From 18S ribosomal sequence data to evolution of morphology among the fungi. *Can. J. Bot. (Suppl. 1)* **73**, S677–S683 (1995).
- Spatafora, J. W. Ascomal evolution of filamentous ascomycetes: evidence from molecular data. *Can. J. Bot. (Suppl. 1)* **73**, S811–S815 (1995).
- Berbee, M. L., Carmean, D. A. & Winka, K. Ribosomal DNA and resolution of branching order among the Ascomycota: how many nucleotides are enough? *Mol. Phy. Evol.* **17**, 337–344 (2000).
- Bhattacharya, D. et al. Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. *Mol. Biol. Evol.* **17**, 1971–1984 (2000).
- Liu, Y. J., Whelen, S. & Hall, B. D. Phylogenetic relationships among ascomycetes: evidence from the RNA polymerase II subunit. *Mol. Biol. Evol.* **16**, 1799–1808 (1999).
- Collins, T. M., Wimberger, P. H. & Naylor, G. J. P. Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol.* **43**, 482–496 (1994).
- Maddison, D. R. Phylogenetic methods for inferring the evolutionary history and process of change in discretely valued characters. *Ann. Rev. Entomol.* **39**, 267–292 (1994).
- Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* **255**, 37–45 (1994).
- Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
- Elix, J. A. in *Lichen Biology* (ed. Nash, T. H.) 154–180 (Cambridge Univ. Press, Cambridge, 1996).
- Rambold, G. & Triebel, D. The inter-lecanoralean associations. *Bibliotheca Lichenologica* **48**, 1–201 (1992).
- Hibbett, D. S., Gilbert, L.-B. & Donoghue, M. J. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature* **407**, 506–508 (2000).
- Lutzoni, F. & Vilgalys, R. *Omphalina* (Basidiomycota, Agaricales) as a model system for the study of coevolution in lichenized fungi. *Crypt. Bot.* **5**, 82–97 (1995).
- Lutzoni, F. Phylogeny of lichen- and non lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. *Syst. Biol.* **46**, 373–406 (1997).
- Lutzoni, F. & Pagel, M. Accelerated evolution as a consequence of transition to mutualism. *Proc. Natl Acad. Sci. USA* **94**, 11422–11427 (1997).
- Kranner, I. & Lutzoni, F. in *Plant Responses to Environmental Stresses: From Phytohormones to Genome Reorganization* (ed. Lerner, H. R.) 591–628 (Marcel Dekker, New York, 1999).
- Wilson, I. & Balding, D. Genealogical inference from microsatellite data. *Genetics* **150**, 499–510 (1998).
- Hillis, D. M., Moritz, C. & Mable, B. K. *Molecular Systematics* 2nd edn (Sinauer, Sunderland, 1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Lewis and K. Pryer for insightful suggestions; J. Spatafora, D. Armaleo, C. and W. Culbertson for initiating the first phase of this project; U. Soehring, J. Spatafora, J. Johnson and S. LaGreca for providing unpublished sequences; J. Belsle for assistance with Figs 1 and 2; J. Crodian for technical assistance; I. Brodo, D. Gernandt, C. Keller, T. Lumbsch, J. Miadlikowska, J. Platt and U. Soehring for providing lichen material or DNA samples; F. Kauff, T. Bjelland, B. Buel, P. M. Jørgensen and M. Schultz for the identification of some lichen specimens; and M. Blackwell for comments on the manuscript. This work was supported by a grant from the US National Science Foundation to F.L. and the Pritzker Foundation Fund of The Field Museum. M.P. is supported by the Leverhulme Trust and the Biotechnology and Biological Sciences Research Council of the UK.

Correspondence and requests for materials should be addressed to F.L. (e-mail: flutzoni@fmmh.org).

Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates

Michael J. Stanhope, Andrei Lupas, Michael J. Italia, Kristin K. Koretke, Craig Volker & James R. Brown

Bioinformatics, GlaxoSmithKline, 1250 South Collegeville Road, UP1345, Collegeville, Pennsylvania 19426, USA

Horizontal gene transfer (HGT) has long been recognized as a principal force in the evolution of genomes¹. Genome sequences of Archaea and Bacteria have revealed the existence of genes whose similarity to loci in distantly related organisms is explained most parsimoniously by HGT events^{2–4}. In most multicellular organisms, such genetic fixation can occur only in the germ line. Therefore, it is notable that the publication of the human genome reports 113 incidents of direct HGT between bacteria

and vertebrates⁵, without any apparent occurrence in evolutionary intermediates, that is, non-vertebrate eukaryotes. Phylogenetic analysis arguably provides the most objective approach for determining the occurrence and directionality of HGT^{6,7}. Here we report a phylogenetic analysis of 28 proposed HGT genes, whose presence in the human genome had been confirmed by polymerase chain reaction (PCR)⁵. The results indicate that most putative HGT genes are present in more anciently derived eukaryotes (many such sequences available in non-vertebrate EST databases) and can be explained in terms of descent through common ancestry. They are, therefore, unlikely to be examples of direct HGT from bacteria to vertebrates.

Most of the phylogenetic analyses (see Methods and Fig. 1 for a description of the phylogenetic principles regarding acceptance or rejection of HGT) for the 28 loci analysed here supported the monophyly of eukaryotes, with a non-vertebrate ('non-vertebrate' refers to invertebrate animals as well as fungi, plants and protists) eukaryote at the base of this clade (Table 1 and Fig. 2). Such outcomes were congruent between both parsimony and distance-based methods of phylogenetic analysis. The explanations for why such a conclusion was not reached previously are somewhat varied. We found that orthologues to various proteins, purported to be HGTs in the recent human genome publication⁵, could actually be

found in the expressed sequence tags (ESTs) of non-vertebrates, which are accessible through the "EST others database" on the NCBI BLAST page (<http://www.ncbi.nlm.nih.gov/BLAST>).

A common search result was the collection of a homologue from the slime mould *Dictyostelium discoideum*, which when included in the alignment resulted in eukaryotic monophyly with *D. discoideum* at the base. One such specific example involves the protein known as formiminotransferase cyclodeaminase (accession number AAG01853.1; this and other accession numbers used here are from Table 24 in ref. 5). This seems to be a duplicate enzyme in eukaryotes, with a cyclodeaminase carboxy-terminal domain and a formiminotransferase amino-terminal domain. The only bacterium at present that can clearly be shown to have this arrangement is *Streptococcus pyogenes*. All other taxa either have separate proteins or just the cyclodeaminase. Alignments constructed on this C-terminal half (chosen because more taxa were available for this enzyme) resulted in phylogenetic trees with 100% bootstrap support for the monophyly of eukaryotes, with *D. discoideum* (EST assembly of AU060687.1 and C94072.1) at the base of the clade (Fig. 2a).

Another example of an orthologue that was missed owing to overlooking the EST others database involves monoamine oxidase (AAA59548.1 and AAB27229.1), cited as an example of a gene

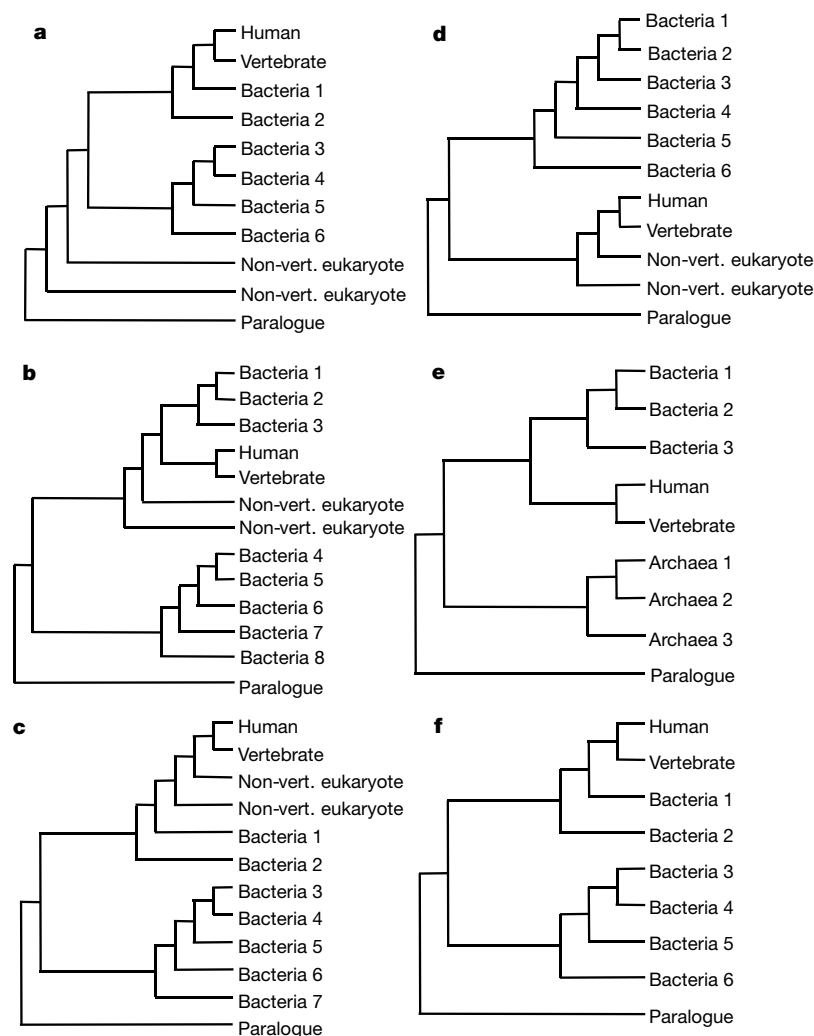


Figure 1 Hypothetical phylogenetic trees showing various evolutionary models relating to acceptance or rejection of HGT. Parologue, an anciently derived paralogous sequence; Non-vert. eukaryote, non-vertebrate eukaryote. **a**, Phylogenetic history in support of bacteria-vertebrate HGT. **b**, Phylogenetic history in support of vertebrate-bacteria HGT.

c-e, Phylogenetic history rejecting any sort of HGT involving bacteria and vertebrates. **f**, Phylogenetic history which, owing to the absence of a non-vertebrate eukaryote, is considered ambiguous with regard to bacteria-vertebrate HGT.

horizontally transferred into vertebrates, which probably conferred a physiological selective advantage and was thus fixed in vertebrate evolution⁵. This gene may indeed have an important neurological function^{8–10}; however, the presence of a *D. discoideum* orthologue at the base of monophyletic eukaryotes (Fig. 2b) argues strongly against the acquisition of this locus by means of HGT.

A different methodological reason for several of the genes in the human genome report⁵ being considered as bacteria–vertebrate HGTs, was that phylogenetics was not the analytical approach, and

that the conclusions were instead derived largely from top BLAST hit results. In several instances the top BLAST hit was indeed a bacterial species, whereas further down the list of significant BLAST hits one finds a non-vertebrate eukaryote. When such sequences were properly aligned, the resulting phylogenetic trees often supported the monophyly of eukaryotes with the non-vertebrate eukaryote at the base. For example, searches against the non-redundant (NR) database involving the short peptide AAF24044.1 resulted in relatively few hits, with a top BLAST hit

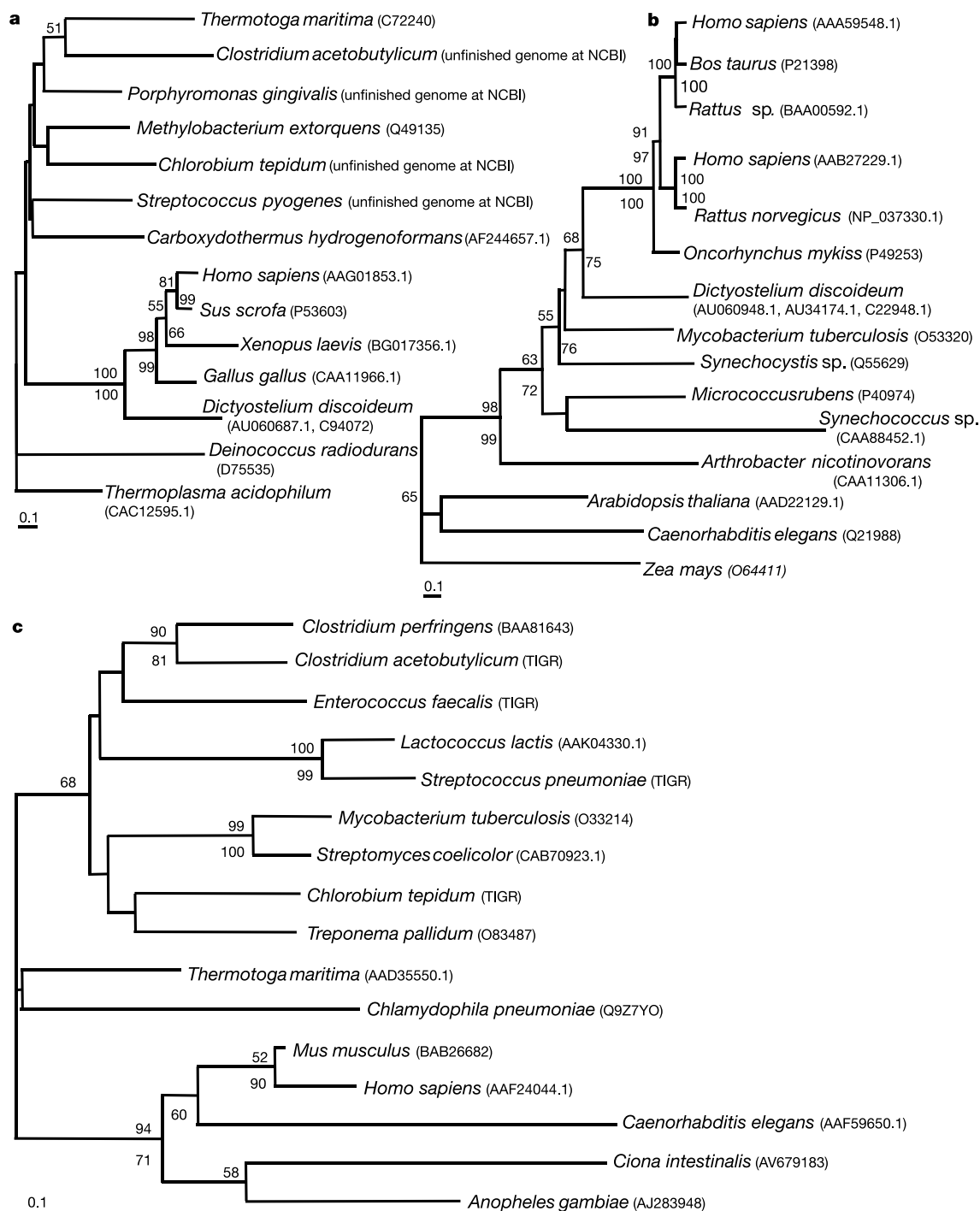


Figure 2 Neighbour-joining phylogenetic trees on the basis of PAM Dayhoff sequence-divergence matrixes, showing evolutionary history of various loci from Table 1. Branch lengths are drawn proportional to the amount of sequence change. Numbers at various nodes indicate bootstrap support values for both neighbour-joining (top) and parsimony

(bottom) analyses; only values in excess of 50% are indicated. **a**, AAG01853.1, formiminotransferase cyclodeaminase. **b**, AAA59548.1, monoamine oxidase A and AAB27229.1, monoamine oxidase B. **c**, AAF24044.1, uncharacterized protein.

outside Mammalia corresponding to the bacterial species *Thermotoga maritima*, and with *Caenorhabditis elegans* listed much further down the list. Despite this arrangement of significant BLAST hits, once the sequences were all aligned and phylogenetic analyses performed, *C. elegans* fell in as the sister group to mammals, and eukaryotes were monophyletic. This conclusion is substantiated further through a search of the EST others database, revealing a sequence for *Anopheles gambiae* (mosquito; AJ283948) and *Ciona intestinalis* (ascidian; AV679183), with the resulting topology again supporting eukaryote monophyly, this time inclusive of two mammals and three invertebrates from disparate phyla (Fig. 2c).

If a particular human gene was either a highly divergent member of a gene family widely found in both vertebrates and non-vertebrates, or short in length but highly matched a motif within a larger bacterial protein, then vertebrates and bacteria were often top hits in BLAST searches. An example is the 222-amino-acid surfactin synthetase domain (IGI_M1_ctg25107_24) for which the top BLAST hit is Srf1, a 3,587-amino-acid protein in *Bacillus subtilis*. Subsequent phylogenetic analyses on the conserved regions between bacteria and humans, as well as fungi and Nematoda, resulted in a monophyletic cluster of eukaryotic sequences. Ribosomal protein S6-glutamic acid ligase (IGI_M1_ctg12741_7) exists in humans, mouse, frog and fly (these latter two were detected in EST others database), and all eukaryotic sequences are monophyletic with high bootstrap support.

One of the twenty-eight loci in question was rejected as a possible bacteria-vertebrate HGT (BAA11432.1) because of its very limited bacterial spectrum. This gene did not strictly adhere to our phylogenetic criteria for rejection, or to the reverse transfer; however, it had very few bacterial orthologues (only *Mycobacterium tuberculosis* RV1834 (Q50599) and *Pseudomonas aeruginosa* PA3429

(C83216)), with a broader representation of vertebrates (fish, frog and a few mammals). The two bacteria did not form a clade in the resulting phylogeny. The most parsimonious explanation of this spectrum and branching arrangement is the evolution of an α/β hydrolase in vertebrates, which was then transferred independently into *Mycobacterium* and *Pseudomonas*. In our opinion this limited spectrum is a compelling argument against bacteria-vertebrate HGT. Such an assortment of taxa would necessitate that this locus evolved in bacteria, was subsequently lost in a host of independent lineages (remembering that there are a significant number of complete bacterial genome sequences available), retained in only *Mycobacterium* and *Pseudomonas*, and then one or another of these taxa transferred this gene into a vertebrate lineage germ line. A similar, although more complex case, is presented by a family of paralogous human transporters (CAB81772.1, AAB59448.1, AAA36608.1 and AAC41747.1; see Table 1), which contains two further members (CAC00574.1, AW904970) listed in the Supplementary Table of ref. 5.

In our analyses of these data there are a few examples where the evolutionary history is unclear (listed under ambiguous in Table 1), and thus for which a possible bacteria-vertebrate transfer cannot be rejected. However, no such phylogenetic criteria exist for these same genes to substantiate the claims of bacteria-vertebrate HGT. Most of our analyses and phylogenetic topologies are highly consistent with the view that vertebrates and bacteria share these loci through common ancestry, involving a succession of non-vertebrate eukaryote intermediates. A further point arising from our analysis is that the evolutionary relationships among proteins cannot be concluded solely from the ranking of database hits in homology searches (for example, BLAST reports). This is not a new conceptual point (see refs 7, 12, 13), but one that seems to have been overlooked in this instance. Phylogenetic analysis must be a central component of any

Table 1 Alternative analysis of proposed vertebrate acquisitions of bacterial genes

Human protein* (accession number)	Phylogenetic support of vert-bac HGT (Fig. 1b)	Phylogenetic rejection of bac-vert HGT (Fig. 1c, d, e)	Other rationale for rejecting bac-vert HGT	Ambiguous: absence of non-vertebrate eukaryote (Fig. 1f)
AAG01853.1		Yes†		
CAB81772.1			Yes‡	
AAB59448.1			Yes‡	
AAA36608.1			Yes‡	
AAC41747.1			Yes‡	
BAA11432.1			Yes§	
CAB59628.1	Yes			
BAA91273.1		Yes†		
CAA75608.1		Yes		
AAA59548.1		Yes†		
AAB27229.1		Yes†		
AAF12736.1		Yes†		
AAA51565.1				Yes
BAA92632.1			Yes	
BAA34458.1		Yes		
AAF24044.1		Yes†		
BAA91839.1				Yes
BAA92073.1			Yes¶	
BAA92133.1				Yes
BAA91174.1				Yes
AAA60043.1		Yes		
BAA86552.1		Yes†		
IGI_M1_ctg12741_7		Yes†		
IGI_M1_ctg13238_61		Yes†		
IGI_M1_ctg13305_116		Yes†		
IGI_M1_ctg14420_10		Yes†		
IGI_M1_ctg16010_18		Yes		
IGI_M1_ctg25107_24		Yes		

For the 28 loci there was no phylogenetic evidence to support the claim of bacteria-vertebrate HGT (Fig. 1a). Sequence BAA34458.1 did not correspond to a β -lactamase-like hydrolase.

* Accession numbers correspond to those given in Table 24 of ref. 5. Sequences IGI_M1_ctg19153_147 and IGI_M1_ctg16227_58 from Table 24 were not found in the IPI_1 database of putative human proteins.

† Searches in which a sequence was recovered from EST others, which proved instrumental in rejecting the bacteria-vertebrate hypothesis.

‡ Large family of paralogous transporters in metazoa; few, phylogenetically diverse bacteria.

§ Few, phylogenetically diverse bacteria.

|| Unique N terminus fused with C-terminal decarboxylase, widely found in Bacteria, Archaea and Eucarya.

¶ Bacterial and human genes are not orthologous.

protein family or genome annotation effort. Importantly, phylogenetic reconstruction is critical to synthesizing, from the growing wealth of sequence data, a more comprehensive view of genome evolution. □

Methods

Data collation and analysis

Our analysis involved a set of 28 loci that had been proposed as instances of bacteria–vertebrate HGT, and for which PCR had been used to verify their presence in the human genome⁵. Each protein sequence was searched using the appropriate BLASTP or TBLASTN¹⁴ algorithm against the following databases downloaded on 19 February 2001 from NCBI: non-redundant proteins; EST others nucleotides (GenBank non-mouse and non-human EST entries); unfinished microbial genome nucleotides; and the complete genome sequences of the nematode *C. elegans* and the insect *Drosophila melanogaster*. ClustalW¹⁵ was used to align the resulting set of significant homologues, and the alignments were then refined by eye using the GCG or GeneDoc sequence editors. In several instances BLAST searches resulted in a large number of homologous sequences, which served as input for a preliminary phylogeny designed to assess paralogy (sequence homology due to gene duplication) and orthology (sequence homology due to speciation). This tree then served as a framework on which to exclude the more distantly related paralogues in a follow-up alignment and phylogeny. All potentially ambiguous gaps (that is, those involving complex insertion or deletion events) were removed before phylogenetic analysis. We coded all remaining gaps as missing data.

The amino-acid sequence alignments were analysed using maximum parsimony and neighbour joining, as implemented in the phylogenetic package PHYLIP¹⁶. Parsimony analyses randomized the input order 20–50 times, depending on the number of sequences in the alignment (that is, greater number of random additions for larger alignments). The distance matrixes that served as input for neighbour-joining analyses were calculated using the point-accepted-mutation (PAM) Dayhoff substitution model. Clade strength was assessed with bootstrap, using 1,000 replications. The alignments are available in Supplementary Information.

Phylogenetic principles for accepting or rejecting HGT

The branching arrangement of the resulting phylogenies is a critical component of an HGT assessment. In the case of a gene transfer from bacteria to vertebrates, the necessary phylogeny would show eukaryote paraphyly with vertebrates separated from non-vertebrates by a paraphyletic (a group that contains some but not all of the descendants of a common ancestor) assemblage of bacterial lineages, where some of the bacteria were more closely related to vertebrates than other bacteria, and in which most (or all) of available bacterial sequences form a clade with vertebrates (Fig. 1a). Horizontal gene transfer of the opposite direction (vertebrate–bacteria) would be expected to have a topology in which one or a few bacteria were sister group to the vertebrates, joined next by non-vertebrate taxa, and the remaining bacteria (if any) were outside this clade (Fig. 1b). Both such trees would need a root, which could come from Archaea or a paralogous gene. The clearest phylogenetic rejection of proposed bacteria–vertebrate HGT would involve eukaryotic monophyly, with a non-vertebrate eukaryote contained within that clade (Fig. 1c, d). Another example of phylogenetic rejection would include the respective monophyly of Archaea, Eucarya and Bacteria (or conversely, just Eucarya and Bacteria) in a tree rooted at a paralogous sequence, but with only vertebrates represented as eukaryotes (that is, bacterial monophyly negates the bacteria–vertebrate HGT hypothesis; Fig. 1e). Several ambiguous cases could arise, the most likely relating to the absence of non-vertebrate orthologous genes. A paraphyletic group of bacteria forming a clade with vertebrates, joined next by either Archaea or a paralogous gene family, could be a consequence of the poor representation of non-vertebrate genomes in contemporary databases, or could reflect an actual HGT from bacteria to vertebrates (Fig. 1f). The absence of a sequence from either a nematode (*C. elegans*) or fruit fly (*D. melanogaster*)—whose nearly complete genomes are available—in such a topology is not sufficient evidence to conclude bacteria–vertebrate HGT. Until a much broader sampling of non-vertebrate genomes are completed these cases should remain ambiguous.

Received 18 April; accepted 24 May 2001.

- Gray, M. W. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* **141**, 233–357 (1992).
- Nelson, K. E. *et al.* Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 (1999).
- Ruepp, A. *et al.* The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513 (2000).
- Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128 (1999).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Smith, M. W., Feng, D.-F. & Doolittle, R. F. Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* **17**, 489–493 (1992).
- Logsdon, J. M. Jr & Faguy, D. M. Evolutionary genomics: *Thermotoga* heats up lateral gene transfer. *Curr. Biol.* **9**, R747–R751 (1999).
- Brunner, H. G., Nelen, M., Breakefield, X. O., Ropers, H. H. & van Oost, B. B. A. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science* **262**, 578–580 (1993).
- Cases, O. *et al.* Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA. *Science* **268**, 1763–1766 (1995).
- Deckert, J. *et al.* Excess of high activity monoamine oxidase A gene promoter alleles in female patients with panic disorder. *Hum. Mol. Genet.* **8**, 621–624 (1999).

- Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote–eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502 (1997).
- Kyrpides, N. C. & Olsen, G. J. Archaeal and bacterial hyperthermophiles—horizontal gene exchange or common ancestry? *Trends Genet.* **15**, 298–299 (1999).
- Eisen, J. A. Phylogenomics—improving functional predictions for uncharacterized genes by evolutionary analysis. *PCR Methods Appl.* **8**, 163–167 (1998).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. (Department of Genetics, University of Washington, Seattle, 2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Correspondence and requests for materials should be addressed to J.R.B. (e-mail: James_R_Brown@sbphrd.com) or M.J.S. (e-mail: Michael_J_Stanhope@sbphrd.com).

Divergent sexual selection enhances reproductive isolation in sticklebacks

Janette Wenrick Boughman

Department of Zoology, University of British Columbia, Vancouver BC V6T 1Z4, Canada

Sexual selection may facilitate speciation because it can cause rapid evolutionary diversification of male mating signals and female preferences. Divergence in these traits can then contribute to reproductive isolation^{1–3}. The sensory drive hypothesis predicts that three mechanisms underlie divergence in sexually selected traits⁴: (1) habitat-specific transmission of male signals^{5–7}; (2) adaptation of female perceptual sensitivity to local ecological conditions⁸; and (3) matching of male signals to female perceptual sensitivity^{4,9}. I test these mechanisms in threespine sticklebacks (*Gasterosteus* spp.) that live in different light environments. Here I show that female perceptual sensitivity to red light varies with the extent of redshift in the light environment, and contributes to divergent preferences. Male nuptial colour varies with environment and is tuned to female perceptual sensitivity. The extent of divergence among populations in both male signal colour and female preference for red is correlated with the extent of reproductive isolation in these recently diverged species. These results demonstrate that divergent sexual selection generated by sensory drive contributes to speciation.

Sexual selection can act through sensory drive to favour different mating signals and preferences in different environments. The sensory drive hypothesis predicts three mechanisms that may cause such divergence. First, females are likely to prefer conspicuous signals. Because habitat influences the physics of signal transmission, relative conspicuousness of signals should vary with habitat⁴. With visual signals, conspicuousness is enhanced when a signal differs from the background light. In relatively clear, fresh water, blue and red are high-contrast signal colours¹⁰. In tea-stained fresh water, red signals are likely to be masked by the redshifted background light, but black should be high contrast¹¹. Second, perceptual sensitivity should vary with habitat^{8,12} and lead to variation in preference. For example, the spectral quality of ambient light is likely to influence colour perception¹². Several sources of selection may favour perceptual systems that work effectively in the local environment, because the visual environment can affect prey and predator detection as well as mate detection^{12,13}. Third, male signals that match female perception are easier to detect and

should be preferred by females^{4,9}. Over evolutionary time, the effect of habitat on signal conspicuousness and female perception can lead to divergent signals, perceptual sensitivity, and even female preference. Initial divergence would occur most readily in allopatric populations but could continue in sympatry. This divergence can contribute to reproductive isolation and ultimately to speciation.

Given that habitats differ in water colour, which affects the transmission of colour signals, I tested the following predictions of sensory drive: that male nuptial colour varies in correlation with this variation in habitat; that female perception also varies with habitat; and that signal variation matches perceptual variation. I then tested two predictions of the 'speciation by sexual selection' hypothesis: that mate preferences diverge, and that divergence in mating signals and preferences contribute to reproductive isolation. I studied six populations of recently diverged threespine sticklebacks from four lakes in coastal British Columbia, Canada: Paxton and Cranby lakes on Texada Island, and Enos and Brannen lakes on Vancouver Island. Two lakes (Paxton and Enos) each contain two reproductively isolated ecotypes of sticklebacks, benthics and limnetics. Benthics feed on macro-invertebrates and mate in the vegetated areas of the littoral zone, whereas limnetics feed on plankton in the open water and mate in the unvegetated areas of the littoral zone at shallower depths than the benthics^{14,15}. The other two lakes (Cranby and Brannen) have a single type of stickleback whose pattern of habitat use is intermediate between the limnetic and benthic ecotypes. All lake sticklebacks have descended from marine ancestors since the Pleistocene epoch. Benthics and limnetics are probably the result of separate freshwater invasions¹⁶ and may have been slightly diverged before first contact. Divergence among populations is so recent that little genetic incompatibility has built up, and postmating isolation is primarily ecological¹⁷.

The nesting habitats of the six populations differ in water colour^{14,15} (Fig. 1), which is probably caused by differences in the amount of decaying organic matter and the extent of vegetative cover. The two limnetics (Paxton and Enos) live in the least redshifted habitat, which is similar to the Paxton benthic habitat ($\lambda p50$ (the median wavelength in a spectral scan) Paxton limnetic = 596 nm; Enos limnetic = 607 nm; Paxton benthic = 611 nm; pooled standard error = 4.9). The two solitary types (Cranby and Brannen) and Enos benthics live in significantly more redshifted habitats ($\lambda p50$ Enos benthic = 620 nm; Cranby = 619 nm; Brannen = 627 nm; pooled standard error = 5.4; least significant difference = 10.3).

I scored two aspects of male nuptial colour: area and intensity of red coloration. Populations differ in red area (Fig. 1a; $F_{5,162} = 166.3$, $P < 0.0001$) and red intensity (Paxton limnetic = 2.5; Enos limnetic = 2.7; Paxton benthic = 2.3; Enos benthic = 0.11; Cranby = 0.92; Brannen = 0.14; $F_{5,160} = 87.1$, $P < 0.0001$). Area and intensity of red coloration correlate with the extent of redshift in background light (rank correlation coefficient: area, $r = -1.0$, $n = 6$, $P < 0.01$ (Fig. 1); intensity, $r = -0.94$, $n = 6$, $P < 0.05$). Thus, males display nuptial colours that are likely to contrast with the water colour of the habitat in which they mate and thus transmit readily, confirming the first prediction of sensory drive.

Using the optomotor response¹⁸ I found differences between populations in female detection thresholds for red light ($F_{5,443} = 32.1$, $P < 0.0001$). A lower threshold implies higher sensitivity. Limnetic females in both Paxton and Enos lakes are more sensitive to red than benthic females ($F_{1,438} = 16.2$, $P < 0.0001$). Females from populations with red males are more sensitive to red light than females from populations with black males ($F_{1,438} = 18.0$, $P < 0.0001$). This perceptual variation correlates with variation in water colour (Fig. 1b; $r = 0.94$, $n = 6$, $P < 0.05$), suggesting that the environment strongly influences perceptual sensitivity, confirming

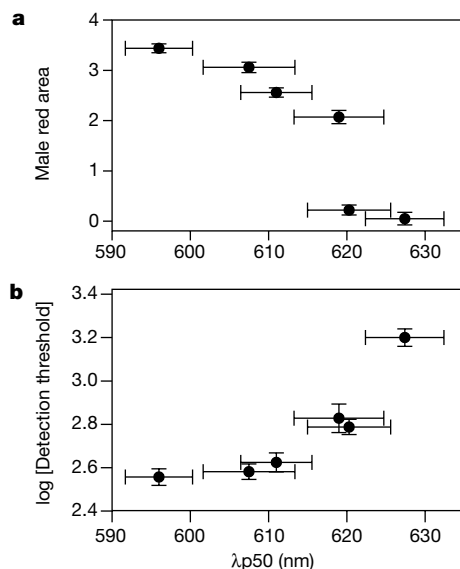


Figure 1 Influence of water colour ($\lambda p50$ values for nesting habitat) on sexually selected traits in males and females. Values given are population means $\pm$ standard error (also for Figs 2 and 3). $\lambda p50$ provides an index of dominant water colour. **a**, Area of red coloration in males correlates negatively to $\lambda p50$ ($r = -1.0$, $P < 0.01$). As light becomes more redshifted, males display less red. Area of red coloration in males is scored from 0 (least red) to 5 (maximal red). **b**, Female detection threshold shows a significant positive correlation to $\lambda p50$ ($r = 0.94$, $P < 0.05$). Females with high detection thresholds (low sensitivity) to red light mate in more redshifted habitats. Female sensitivity to red is high when the detection threshold is low and is measured in log photon flux (in units of $\mu\text{mol photons m}^{-2} \text{s}^{-1}$).

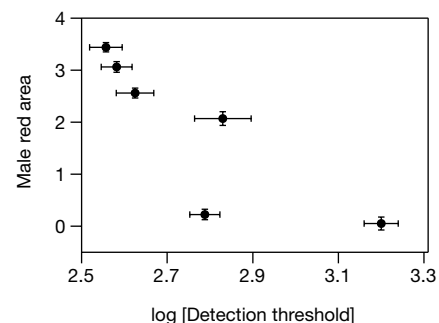


Figure 2 Relationship between female detection threshold for red and area of red coloration in males ($r = -0.94$, $P < 0.05$). Female sensitivity to red and area of red coloration in males measured as in Fig. 1.

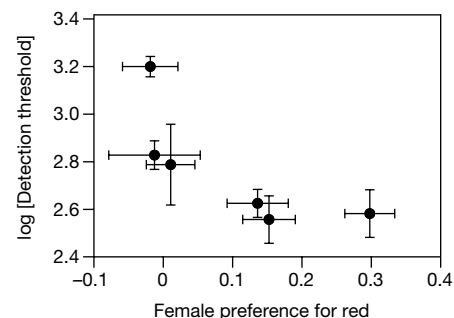


Figure 3 Relationship between female perception of red light and preference for red signals ($r = -0.94$, $P < 0.05$).

the second prediction of sensory drive.

Female sensitivity to red light correlates significantly with the area of red coloration in males (Fig. 2; $r = -0.94$, $n = 6$, $P < 0.05$), and non-significantly to red intensity ($r = -0.77$, $n = 6$, $P < 0.10$). Thus, male signals seem to be fine-tuned to female perceptual sensitivity, confirming the third prediction of sensory drive. Signal matching is also predicted by the pre-existing bias hypothesis⁹. Covariation could be caused by male signals evolving to match a pre-existing bias⁹, or could arise because both traits evolved in response to a shared environment⁴; therefore, causality remains to be determined.

A critical prediction of speciation by sexual selection is that mate signals and preferences diverge and this divergence contributes to reproductive isolation. In behavioural trials, I estimated female preference functions for red nuptial colour in the six populations by measuring the number of times females examined a male's nest per min, and measured reproductive isolation as the occurrence of spawning between males and females from different populations. I used analysis of covariance (ANCOVA) to estimate preference for red males by calculating the slope of the linear regression of preference (the log of the rate of nest examinations) on area of red coloration (slope estimates: Paxton limnetic = 0.15; Enos limnetic = 0.30; Paxton benthic = 0.14; Enos benthic = 0.01; Cranby = -0.01; Brannen = -0.02; slope heterogeneity $F_{5,135} = 2.3$, $P < 0.05$). Limnetics from both lakes differ significantly from other populations ($F_{1,135} = 53.3$, $P < 0.001$) and black populations differ significantly from red ones ($F_{1,135} = 7.9$, $P < 0.01$). Thus, the strength of female preference for red signals has diverged among populations. Divergence in preference for red depends, in part, on perceptual sensitivity to red (Fig. 3; $r = -0.94$, $n = 6$, $P < 0.05$).

Next, I used logistic regression to estimate spawning rates between each pair of populations, correcting for differences between populations in male propensity to spawn. I corrected all population trait means for phylogenetic similarity¹⁹. I then analysed the relationship between the divergence in sexually selected traits (male signals and female preferences) and reproductive isolation using generalized least squares (GLS)¹⁹ and Mantel tests, using the six populations as replicates.

The divergence I found in both male signal colour and female

preference for red are significantly correlated with the extent of reproductive isolation between populations (Fig. 4). The extent of divergence in male signals is negatively correlated with the amount of reproductive isolation by both GLS ($b = -0.11$, $n = 6$, $P < 0.001$) and Mantel ($r = -0.93$, $P < 0.03$) tests. I found a similar pattern for divergence in female preference and reproductive isolation with both GLS ($b = -0.87$, $n = 6$, $P < 0.05$) and Mantel ($r = -0.90$, $P < 0.05$) tests. Thus, greater divergence in male signals and female preference contributes to greater reproductive isolation between populations.

Body size is also known to contribute to reproductive isolation between ecotypes²⁰, but differences in body size do not account for the pattern shown in Fig. 4. Paxton benthics and limnetics differ most in body size²⁰ and should have the lowest spawning rate; however, they spawn at nearly twice the rate of Enos benthics and limnetics (Paxton benthics and limnetics = 0.30; Enos benthics and limnetics = 0.17). Results are consistent with sensory drive predictions, but both Fisherian runaway^{1,2} and condition-dependent²¹ hypotheses predict a correlation between male signal and female preference and could contribute to this correlation. Neither hypothesis makes specific predictions about female perception. However, the signalling environment strongly influences both traits, which is only indirectly predicted for male colour by the Fisher process if the strength of natural selection opposing mate choice is negatively correlated with water colour. Enhancing conspicuousness in the local environment is likely to increase predation risk, so runaway selection should produce a pattern opposite to the one observed here. Condition dependence predicts the relationship between male colour and environment if the cost-benefit relationship of signalling varies closely with water colour, or if carotenoid availability correlates negatively with water colour. A model that considers the interaction between condition dependence and perception predicts the pattern found; however, this prediction arises from effects of environment on signal detectability²² and thus invokes a mechanism similar to sensory drive.

Perception, preferences and signals diverge in correlation with habitat, implying that some habitat specialization may precede or coincide with their divergence. Ecological diversification and repro-

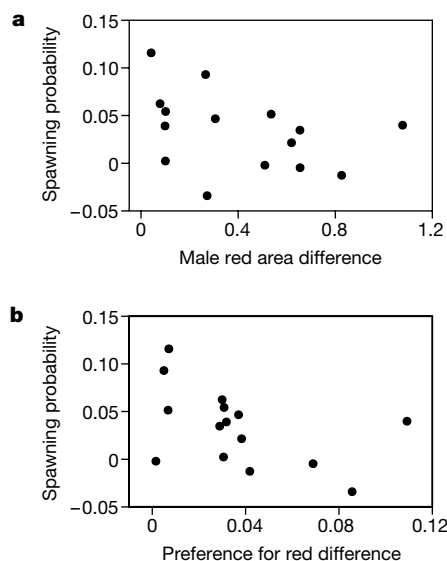


Figure 4 Relationship between the divergence among populations in male signal colour and female preference for red, and reproductive isolation between populations. Values given are for pairs of populations and are corrected for phylogeny. **a**, Difference in area of red coloration in males correlates negatively with spawning probability. **b**, Difference in

strength of female preference for red signals correlates negatively with spawning probability. For both traits, as the extent of divergence increases the probability of mating decreases significantly with both GLS and Mantel analyses. See text for details.

ductive isolation are probably both necessary for the persistence of species with currently parapatric or sympatric distributions, and their joint occurrence is a feature of several models of speciation^{23,24}. Assortative mating on the basis of a trait correlated with resource specialization, such as body size, also occurs in these stickleback populations²⁰. The combined forces of divergent sexual selection and competition for resources in sticklebacks²⁵ are expected to facilitate the rapid evolution of phenotypic differences and speciation²⁶.

Closely related species often differ markedly in sexually selected traits. Understanding the role sexual selection has in speciation requires that we identify both the causes of diversity in sexually selected traits and the consequences on pre-mating isolation. My results support important predictions of the hypothesis that sexual selection facilitates speciation, and provide evidence that sensory drive is a critical selective mechanism. The environment is implicated: when local habitats differ, divergence in sexually selected traits is a likely result. Thus, sensory drive provides a general mechanism for divergence in signalling systems, and could figure prominently in the speciation process. □

Methods

Measuring ambient light

I measured spectral radiance in nesting habitats of the six populations with a spectroradiometer (Licor underwater (UW1800) or Ocean Optics (S2000)). At two sites in each habitat where males were observed nesting I recorded sidewelling light (perpendicular to the water surface) at the approximate nesting depth (0.3–2 m). I calculated $\lambda p50$ values from spectral radiance curves by integrating the area under the curves and taking the median value²⁷. I analysed $\lambda p50$ values with an analysis of variance (ANOVA) test.

Measuring male colour

Just before conducting trials I scored the area and intensity of red on each male used in mating trials according to a six-point scale from 0 (no red area; no red intensity) to 5 (maximal red area; maximal intensity). The maximal area of red coloration included colour on the male's lips, throat, gill covers, and ventral surface (extending back to the anal spines). I used single-degree-of-freedom contrasts to test for differences between populations in male colour. Sample sizes: Paxton limnetic = 34; Enos limnetic = 27; Paxton benthic = 31; Enos benthic = 26; Cranby = 22; Brannen = 23; total = 163.

Measuring female perception

I used a behavioural psychophysical technique relying on the optomotor response¹⁸ to assay female perceptual sensitivity to red light. I used a standard slide projector fitted with a rotating spoked wheel to create a moving visual field; a narrow band 10-nm interference filter (640nm; PTR Optics) to control wavelength; and neutral density filters (PTR Optics) to control light intensity. Fish were light-adapted under a 60-W broad-spectrum bulb (Bulbrite Industries true daylight) for 20 min before testing. My response criterion was that the fish follow the rotation of stripes at the rotational speed (6° s^{-1}). I used a descending method of limits procedure²⁸ with 0.3 log unit decrements and estimated the detection threshold as the light intensity one step above that at which the fish ceased responding to the stimulus. After the fish stopped responding I increased light intensity by one step to determine whether the fish would resume response. I then ascertained this light intensity value with a LiCor 1800 spectroradiometer to determine the absolute threshold values (log of light intensity in units of $\mu\text{mole photons m}^{-2} \text{ s}^{-1}$). I used single-degree-of-freedom contrasts to test for differences between ecotypes and between populations with black or red males. Sample sizes: Paxton limnetic = 82; Enos limnetic = 92; Paxton benthic = 61; Enos benthic = 99; Cranby = 54; Brannen = 61; total = 449.

Measuring female preference

I used no-choice mating trials to assess female preferences for male colour. After a single male had built a nest in a 100-l aquarium I introduced a single gravid female to the aquarium (see refs 20, 29 for details). I scored male colour before each trial. During the 20-min trials I collected data on female response to male courtship and the occurrence of spawning. After each trial I measured female perceptual sensitivity, body size and weight, and verified that the female was ready to spawn. I included only those trials where females were receptive. I estimated female preference functions by ANCOVA with area of red coloration as the covariate and female population as a categorical effect. I calculated the linear regression coefficient of rate of nest examination on area of red coloration for each female population on the basis of trials conducted with males from the females' own population. These slope estimates indicate the strength of preference for male nuptial colour for each female population. Sample sizes: Paxton limnetic = 29; Enos limnetic = 17; Paxton benthic = 23; Enos benthic = 23; Cranby = 26; Brannen = 34; total = 152.

Reproductive isolation and divergence

I used no-choice mating trials, as above, to measure reproductive isolation, pairing males

and females from different populations (for example, Paxton limnetic $\times$ Paxton benthic; Enos benthic $\times$ Brannen). Total sample size was 654. I used logistic regression to estimate spawning rates between each pair of populations, correcting for differences between populations in male propensity to spawn. Next, I estimated the magnitude of divergence in sexually selected traits by calculating the difference between pairs of populations in male signal (red area) and strength of preference for red (slope estimates). This yields 15 pairwise comparisons. I corrected all values for phylogenetic similarity. To do this, I calculated the inverse of the sum of the branch lengths between each pair of populations from a phylogenetic tree³⁰, and used these values as weighting factors. I then analysed the relationship between the divergence in sexually selected traits (male signals and female preferences) and reproductive isolation using GLS¹⁹ and Mantel tests. I used the six populations as replicates.

Received 29 January; accepted 3 April 2001.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Lande, R. Models of speciation by sexual selection on polygenic traits. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725 (1981).
3. West-Eberhard, M. J. Sexual selection, social competition, and speciation. *Q. Rev. Biol.* **58**, 155–183 (1983).
4. Endler, J. A. Signals, signal conditions, and the direction of evolution. *Am. Nat.* **139**, 125–153 (1992).
5. Ryan, M. J. & Wilczynski, W. Evolution of intraspecific variation in the advertisement call of a cricket frog (*Acris crepitans*, Hylidae). *Biol. J. Linn. Soc.* **44**, 249–271 (1991).
6. Marchetti, K. Dark habitats and bright birds illustrate the role of the environment in species divergence. *Nature* **362**, 149–152 (1993).
7. Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).
8. Levine, J. S. & MacNichol, E. F. Jr Visual pigments in teleost fishes: effects of habitat, microhabitat, and behavior on visual system evolution. *Sensory Process.* **3**, 95–131 (1979).
9. Ryan, M. Sexual selection, sensory systems, and sensory exploitation. *Oxf. Surv. Evol. Biol.* **7**, 156–195 (1990).
10. McFarland, W. N. & Munz, F. W. Part III: The evolution of photopic visual pigments in fishes. *Vision Res.* **12**, 1071–1080 (1975).
11. Reimchen, T. E. Loss of nuptial color in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution* **43**, 450–460 (1989).
12. Lythgoe, J. N. *The Ecology of Vision* (Oxford Univ. Press, 1979).
13. Endler, J. A. Variation in the appearance of guppy color patterns to guppies and their predators under different visual conditions. *Vision Res.* **31**, 587–608 (1991).
14. McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): morphological and genetic evidence for a species pair in Enos Lake, British Columbia. *Can. J. Zool.* **62**, 1402–1408 (1984).
15. McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): evidence for a species pair in Paxton Lake, Texada Island, British Columbia. *Can. J. Zool.* **70**, 361–369 (1992).
16. McPhail, J. D. in *The Evolutionary Biology of the Threespine Stickleback* (eds Bell, M. A. & Foster, S. A.) 399–437 (Oxford Univ. Press, 1994).
17. Schluter, D. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) 114–129 (Oxford Univ. Press, 1998).
18. Cronly-Dillon, J. & Sharma, S. C. Effect of season and sex on the photopic spectral sensitivity of the three-spined stickleback *J. Exp. Biol.* **49**, 679–687 (1968).
19. Martins, E. P. & Hansen, T. F. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646–667 (1997).
20. Nagel, L. & Schluter, D. Body size, natural selection, and speciation in sticklebacks. *Evolution* **52**, 209–218 (1998).
21. Kodric-Brown, A. & Brown, J. H. Truth in advertising: the kinds of traits favored by sexual selection. *Am. Nat.* **124**, 309–323 (1984).
22. Schluter, D. & Price, T. D. Honesty, perception, and population divergence in sexually selected traits. *Proc. R. Soc. Lond. B* **253**, 117–122 (1993).
23. Lande, R. & Kirkpatrick, M. Ecological speciation by sexual selection. *J. Theor. Biol.* **133**, 85–98 (1988).
24. Higashi, M., Takimoto, G. & Yamamura, N. Sympatric speciation by sexual selection. *Nature* **402**, 523–526 (1999).
25. Schluter, D. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798–801 (1994).
26. Podos, J. Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* **409**, 185–188 (2001).
27. McDonald, C. G. & Hawryshyn, C. W. Intraspecific variation of spectral sensitivity in threespine stickleback (*Gasterosteus aculeatus*) from different photic regimes. *J. Comp. Physiol. A* **176**, 255–260 (1995).
28. Klump, G. M., Dooling, R. J., Fay, R. R. & Stebbins, W. C. *Methods in Comparative Psychoacoustics* (Birkhauser, Basel, Switzerland, 1995).
29. Rundle, H. D., Nagel, L. M., Boughman, J. W. & Schluter, D. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308 (2000).
30. Taylor, E. B. & McPhail, J. D. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proc. R. Soc. Lond. B* **267**, 2375–2384 (2000).

Acknowledgements

I would like to thank S. Morgan and B. Harvey for help with data collection, and C. Hawryshyn (and his NSERC equipment grant) and P. Jolliffe for the loan of equipment. I also thank J. Endler, S. Kuhnholz, T. Lenormand, S. Otto, A. Poon, H. Rundle, G. Saxer, D. Schluter, J. Smith, S. Vamori and M. Whitlock for comments on the manuscript; Holnam West Materials for access to Paxton Lake; and Fairwinds Corporation for access to

Enos Lake. This work was supported by NSF-NATO and NSF-International Research fellowships to J.W.B. and NSERC operating grants to D. Schluter.

Correspondence and requests for materials should be addressed to J.W.B. (e-mail: boughman@zoology.ubc.ca).

Nodulation of legumes by members of the β -subclass of Proteobacteria

Lionel Moulin, Antonio Munive, Bernard Dreyfus & Catherine Boivin-Masson

Laboratoire des Symbioses Tropicales et Méditerranéennes, IRD-INRA-CIRAD-ENSAM Baillarguet, PB 5035, 34398 Montpellier Cedex 5, France

Members of the Leguminosae form the largest plant family on Earth, with around 18,000 species. The success of legumes can largely be attributed to their ability to form a nitrogen-fixing symbiosis with specific bacteria known as rhizobia, manifested by the development of nodules on the plant roots in which the bacteria fix atmospheric nitrogen, a major contributor to the global nitrogen cycle. Rhizobia described so far belong exclusively to the α -subclass of Proteobacteria, where they are distributed in four distinct phylogenetic branches^{1,2}. Although nitrogen-fixing bacteria exist in other proteobacterial subclasses, for example *Herbaspirillum* and *Azoarcus* from the phylogenetically distant β -subclass, none has been found to harbour the *nod* genes essential for establishing rhizobial symbiosis^{3,4}. Here we report the identification of proteobacteria from the β -subclass that nodulate legumes. This finding shows that the ability to establish a symbiosis with legumes is more widespread in bacteria than anticipated to date.

Recent taxonomic classifications portray rhizobia as belonging to three different branches of the α -subclass of Proteobacteria: the *Mesorhizobium*-*Sinorhizobium*-*Rhizobium* branch, the *Bradyrhizobium* branch and the *Azorhizobium* branch^{1,5}. We recently described a fourth rhizobial branch within the α -subclass of Proteobacteria, containing methylotrophic rhizobial *Methylobacterium*². To explore further the phylogenetic diversity of nitrogen-fixing legume symbionts, we characterized a collection of rhizobia isolated from tropical legumes. We found here that two bacteria isolated from nodules of *Aspalathus* and *Machaerium*

plants were very distant from known rhizobia.

Strain STM678 was originally isolated from the South African legume *Aspalathus carnosa*, which was thought to be nodulated by bacteria of the *Bradyrhizobium* genus⁶. However, we performed phylogenetic analysis of gene sequences of the small subunit of ribosomal RNA (16S rRNA), and found that strain STM678 does not belong to any of the four branches of rhizobia described so far, nor even to the α -subclass of Proteobacteria, but instead belongs to the β -subclass of Proteobacteria (Fig. 1). From this analysis, we found the most closely-related sequences to that of strain STM678 (AJ302311) to be those of *Burkholderia kururiensis* (96.9% identity), *B. brasilense* (96.8% identity) and *B. graminis* (96.8% identity). Phylogenetic analyses of partial sequences of the 23S rRNA gene (AJ302313) and the *dnaK* gene encoding the chaperon heat shock protein (AJ302314) were consistent with the 16S rRNA analysis, thus unambiguously positioning strain STM678 in the *Burkholderia* genus within the β -subdivision of Proteobacteria.

To ensure that the *Burkholderia* strain STM678 was indeed a rhizobium, we checked its ability to re-nodulate a leguminous plant. Because seeds of the original host plant, *A. carnosa*, were not available, we selected as test plant *Macroptilium atropurpureum*, a tropical legume capable of establishing a symbiosis with diverse rhizobia. Over a three-week period, strain STM678 formed 5 to 20 nodules per plant on the roots of *M. atropurpureum* (Fig. 2). The nodules displayed the classical determinate nodule structure, with a central 'infected' tissue containing cells with intracellular bacteria and a peripheral tissue with vascular bundles (Fig. 2). Single colonies re-isolated from surface-sterilized nodules exhibited the characteristics of strain STM678, as assessed by 16S rDNA sequencing and *nodA* analysis using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP; see below). Hence, Koch's postulates were verified. To eliminate the possibility that strain STM678 is a mixture of two different bacteria, a *Burkholderia* and a rhizobium, we isolated spontaneous mutants resistant to chloramphenicol, rifampicin and streptomycin, and showed by 16S rDNA and *nodA* PCR-RFLP analyses that the individual mutants retained the characteristics of both *Burkholderia* and rhizobia. Nodules induced on *M. atropurpureum* by strain STM678 were ineffective in terms of nitrogen fixation, probably because *M. atropurpureum* is not the original symbiotic partner of strain STM678. Supporting this conjecture, strain STM678 was indeed found to contain the *nifH* gene encoding dinitrogenase reductase, a key enzyme in nitrogen fixation (AJ302315). The highest identity values with other *nifH* genes were 81.2% (the α -Proteobacterium *Azorhizobium caulinodans*) and 81.1% (the β -Proteobacterium *Herbaspirillum seropedicae*).

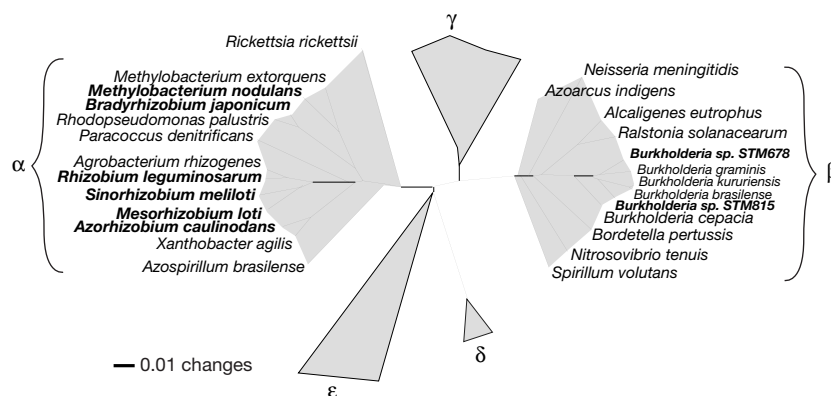


Figure 1 Unrooted 16S rDNA tree of Proteobacteria (purple bacteria). The figure shows the phylogenetic relationships between the different rhizobial genera—as represented by type species in bold—including the new rhizobial *Burkholderia* sp. strains. α , β , δ , γ and ϵ represent the different subdivisions of the Proteobacteria. The tree was constructed by

using the neighbour-joining method and adapted from ref. 5. 16S rDNA sequences of published bacteria are available in GenBank. 16S rDNA from *Burkholderia* sp. STM 678 and *Burkholderia* sp. STM 815 are given in the text (AJ 302311 and AJ 302312).

Nodulation of legumes by rhizobia is controlled by a set of bacterial nodulation (*nod*) genes involved in the production of lipo-chitoooligosaccharides (Nod factors) that act as signalling molecules for nodulating specific legume hosts^{3,4}. The *nodABC* genes are responsible for the synthesis of the core structure of the Nod factor^{7,8}, and as such are present in all rhizobia. We thus looked for the presence of *nodABC* genes in the nodulating *Burkholderia* strain STM678 by PCR amplification (see Methods). Sequencing of the amplified DNA revealed a genetic organisation of *nodAB* genes similar to that found in other rhizobia; that is, with *nodAB* in the same orientation and overlapping and preceded by a NodD-dependent regulatory sequence (*nod* box). A *nodC*-like sequence was found immediately downstream of *nodB*. However this sequence is unlikely to correspond to a functional *nodC* gene as it lacks the ~600-base-pair 5' end of known *nodC* genes. We obtain evidence for the presence elsewhere in the STM678 genome of a longer *nodC* sequence that probably corresponds to the functional *nodC* gene (AJ306730). Such genetic unlinkage of *nodABC* genes in rhizobia is not unprecedented⁹. The sequences of strain STM678 *nodAB* genes (AJ302321) revealed very high similarities with rhizobial Nod protein sequences available in databases, with values ranging from 62.8% (*Sinorhizobium meliloti*) to 77.6% (*Methylobacterium nodulans*) for NodA and from 55.6% (*Rhizobium galegae*) to 70.9% (*Mesorhizobium* sp. N33) for NodB. To examine whether these genes were functional, we constructed a *nodA* mutant by introducing a

lacZ-kanamycin-resistance cassette into the *nodA* gene of strain STM678. The *nodA* mutant did not form any nodules after inoculation on *M. atropurpureum*, even after 30 days, indicating that the *nod* genes that we disrupted are required for nodulation of the *Burkholderia* sp. strain STM678.

By screening among bacteria isolated from root nodules collected from various legumes in French Guiana, we found a second *Burkholderia* rhizobium, strain STM815, isolated from the legume *Machaerium lunatum*. 16S rDNA sequencing (AJ302312) of this strain revealed the following sequence identities with its closest phylogenetic neighbours: 96.9% (*Burkholderia kururienensis*), 96.8% (*B. brasiliense*), 96.6% (*B. graminis*) and 96.9% (strain STM678). These data clearly show that strain STM815 belongs to the *Burkholderia* genus, and most probably to a different species to strain STM678. The nodulation ability of STM815 was confirmed, as described for strain STM678, by inoculation of *M. atropurpureum* in axenic conditions and by re-isolation and characterization of the bacteria isolated from the induced nodules.

The β -subdivision of Proteobacteria contains many bacteria that interact with eukaryotes, including human pathogens, such as *Neisseria* and *Bordetella*, and plant-associated bacteria. These latter bacteria include pathogenic *Ralstonia solanacearum*, rhizospheric *Burkholderia* and endophytic *Azoarcus*. However the β -Proteobacteria had not been reported to include rhizobia, bacteria capable of nodulating leguminous plants. Here, we have identified two rhizobia belonging to the *Burkholderia* genus. These bacteria were isolated in different continents, from legumes belonging to different Papilionoideae tribes, and probably correspond to two distinct species. We have shown that the genetic control of nodulation by the *Burkholderia* sp. strain STM678 involves *nod* genes. Moreover, this strain has been shown to produce Nod factors¹⁰. Hence rhizobia from both the α - and β -Proteobacteria (now termed α - and β -rhizobia) use the same strategy for establishing symbioses with legumes.

Furthermore, we have performed phylogenetic analyses that indicate a much smaller phylogenetic distance between the *nodAB* genes of strain STM678 and other rhizobia (Fig. 3) than between the 16S rRNA genes of α - and β -Proteobacteria (Fig. 1). This suggests that the presence of *nod* genes in both α - and β -rhizobia probably occurred through horizontal gene transfer. This transfer may have occurred after the appearance of legumes on Earth, about 70 million

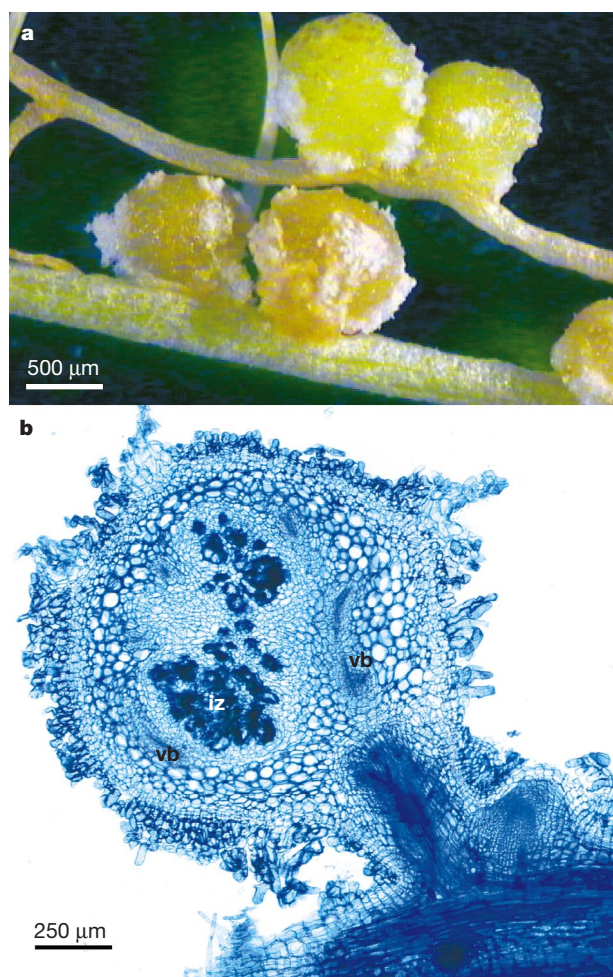


Figure 2 Nodules of *Macroptilium atropurpureum*, three weeks after root inoculation with *Burkholderia* sp. strain STM678. **a**, Root segments with nodules; **b**, longitudinal section showing the typical structure of a determinate nodule with a central zone containing infected cells (iz), and a peripheral region with vascular bundles (vb).

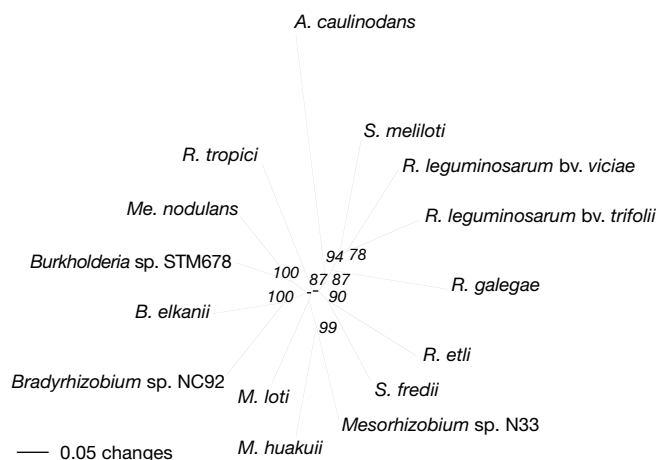


Figure 3 Unrooted NodA tree showing the close phylogenetic relationship between the NodA of strain STM678 and those of α -rhizobia. The tree is based on full-length sequences, and constructed by using the neighbour-joining method. Bootstrap values (% from 1,000 replications) are indicated. NodA sequences of published rhizobia are available in GenBank. NodA from *Burkholderia* sp. STM 678 is given in the text (AJ 302321). A, *Azorhizobium*, B., *Bradyrhizobium*, M, *Mesorhizobium*, Me, *Methylobacterium*, R, *Rhizobium*, S, *Sinorhizobium*.

years ago¹¹. Although rhizobia have been studied for more than 100 years, symbionts of less than 10% of the 750 legume genera have been fully characterized. Our work suggests that characterization of the symbionts of the yet unexplored legumes may reveal the rhizobial nature of additional members of the β -Proteobacteria and possibly other taxonomic classes. Such a study may contribute significantly to the understanding of the origin, and the evolution of, the legume–rhizobia symbioses, and may open new perspectives for engineering beneficial associations. □

Methods

Strains and culture conditions

Strain STM678 was provided by H. P. Spaink (Univ. Leiden). Cells were maintained and grown on yeast extract–mannitol medium².

DNA amplification, sequencing and analysis

Nearly full-length 16S rDNA was amplified and sequenced as previously described². 16S rDNA PCR-RFLP analysis was performed as described² except that *Cfr131*, *HinfI* and *MspI* were used. A 870-bp part of *dnaK* was amplified and sequenced using the primers 5'-GAMGTCAARCBATCATCAA-3' and 5'-TGTCYTTGVCBSMNACRTGCAG-3'. A 370-bp fragment of 23S rRNA gene was amplified and sequenced using the universal primers 5'-AGAGGCGATGAAGGACGT-3' and 5'-ACCTTTCCCTCACGGTACT-3'. A 1,509-bp fragment containing partial *nifH* and *nifD* genes were amplified using the primers 5'-GCCWCTCTAYGGNAARGGNGG-3' and 5'-ATCAGGCCGATCGGGCATT-3' and further sequenced. A 2.6-kb fragment containing the *nodAB* genes of strain STM678 was amplified using the primers 5'-CAGATCNAGDCCBTGAAACGCA-3' (located at the end of *nodD* in rhizobia) and 5'-CTNCGNGCCCARCGNAGTTG-3' (located within *nodC* in rhizobia). This fragment was further sequenced using pairs of degenerate primers defined from conserved motifs of *nodA*, *nodB*, *nodC*, and *nod* box. Two 1-kb overlapping fragments containing part of the *nodC* and *nodI* genes of strain STM678 were amplified using the primer pairs 5'-TAYRTGGTYGAYGACGWTC-3'/5'-CCATACGCACCGTGGTGCTCTTGC-3' and 5'-GGTATCGGACCGAGTACG/5'-TCTTCCATVAWRTGVGTNGTCA-3' (forward primer located at the beginning of *nodC* in rhizobia, reverse primer located within *nodI* in rhizobia) and further sequenced. *nodA* PCR-RFLP analysis was performed on a 455-bp PCR product obtained with the primer pair 5'-TCACARCTCKGGCCGTTCCG-3'/5'-TGGGCSGGNGCAGRCBGA-3' and digested with *Cfr131*, *HinfI* and *HaeIII*. Multiple alignments were performed with CLUSTALX¹². Phylogenetic analyses used the neighbour-joining method and the programmes in PAUP version 4.0b5¹³.

Construction of a *nodA* mutant

A 2.5-kb *XhoI*–*XhoI* fragment containing the *nodAB* genes of strain STM678 was obtained by PCR amplification using modified primers containing additional *XhoI* sites and cloned in the *Sall* site of pJQ200mp18 suicide vector¹⁴. The 4.7-kb *Sall* *lacZ*–kanamycin-resistance cassette of pKOK5¹⁵ was inserted at the *Sall* site of the *nodA* gene cloned in pJQ200mp18. The pJQ200 derivatives obtained, which encoded a counterselectable *sacB* marker, were transformed into *Escherichia coli* XLII, and further introduced by conjugation into a spontaneous chloramphenicol-resistant derivative of strain STM678. Transconjugant colonies grown on YM medium containing 50 $\mu\text{g ml}^{-1}$ kanamycin and 100 $\mu\text{g ml}^{-1}$ chloramphenicol were plated onto YM medium containing 7% sucrose and kanamycin. Sucrose-resistant colonies were screened by PCR to ensure replacement of the wild-type *nodA* gene by the *nodA::lacZ* mutated gene. The *Burkholderia* status of the mutated strain was assessed by 16S rDNA PCR-RFLP.

Plant tests

Seeds were surface-sterilized with concentrated sulphuric acid for 5 min. Plant cultivation and nodulation tests were carried out as described². Effectiveness was estimated by visual observation of plant vigour and foliage colour of 30-day-old plants. Sections were made using a Leica VT1000S Vibratome, and examined after staining with 0.01% methylene blue.

Received 22 January; accepted 2 May 2001.

- Young, J. P. W. & Haukka, K. E. Diversity and phylogeny of rhizobia. *New Phytol.* **133**, 87–94 (1996).
- Sy, A. *et al.* Methylobacterium nodulate and fix nitrogen in symbiosis with legumes. *J. Bacteriol.* **183**, 214–220 (2001).
- Lerouge, P. *et al.* Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781–784 (1990).
- Spaink, H. P. *et al.* A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* **354**, 125–130 (1991).
- van Berkum, P. & Eartly, B. D. in *The Rhizobiaceae* (eds Spaink, H. P., Kondorosi, A. & Hooykaas, P. J. J.) 1–24 (Kluwer Academic, Dordrecht, 1998).
- Deschodt, C. C. & Strijdom, B. W. Effective nodulation of *Aspalathus linearis* spp. *linearis* by rhizobia from other *Aspalathus* species. *Phytophyllactica* **8**, 103–104 (1976).
- Dénarié, J., Debéllé, F. & Promé, J. C. Rhizobium lipo-chito-oligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* **65**, 503–535 (1996).

- Perret, X., Staehelin, C. & Broughton, W. J. Molecular basis of symbiotic promiscuity. *Microbiol. Mol. Biol. Rev.* **64**, 180–201 (2000).
- Zhang, X. X. *et al.* The common nodulation genes of *Astragalus sinicus* rhizobia are conserved despite chromosomal diversity. *Appl. Environ. Microbiol.* **66**, 2988–2995 (2000).
- Boone, C., Olsthoorn, M. M. A., Dakora, F. D., Spaink, H. P. & Thomas-Oates, J. Structural characterisation of lipo-chitin oligosaccharides isolated from *Bradyrhizobium aspalati*, microsymbionts of commercially important South African legumes. *Carbohydr. Res.* **317**, 155–163 (1999).
- Polhill, R. M., Raven, P. H. & Stirton, C. H. in *Advances in Legume Systematics Part 1* (eds Polhill, R. M. & Raven, P. H.) 1–26 (Royal Botanic Gardens, Kew, 1981).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
- Swofford, D. L. PAUP. Phylogenetic Analysis Using Parsimony (and other Methods) Version 4 (Sinauer Associates, Sunderland, Massachusetts, 1998).
- Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* **127**, 15–21 (1993).
- Kokotek, W. & Lotz, W. Construction of a *lacZ*–kanamycin-resistance cassette, useful for site-directed mutagenesis and as a promoter probe. *Gene* **84**, 467–471 (1989).

Acknowledgements

We thank H. P. Spaink for providing strain STM678, M. Neyra for providing the 23S rRNA primers, Y. Prin for help in microscopy studies and C. Huttel for sending plant material. We also thank J. Cullimore and J. Batut for comments and suggestions.

Correspondence and requests for materials should be addressed to C.B.-M. (e-mail: Catherine.Boivin@mpl.ird.fr). EMBL accession numbers: AJ302311 to AJ302315, AJ302321, AJ306730.

Spatial awareness is a function of the temporal not the posterior parietal lobe

Hans-Otto Karnath, Susanne Ferber & Marc Himmelbach

Department of Cognitive Neurology, University of Tübingen, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

Our current understanding of spatial behaviour and parietal lobe function is largely based on the belief that spatial neglect in humans (a lack of awareness of space on the side of the body contralateral to a brain injury) is typically associated with lesions of the posterior parietal lobe. However, in monkeys, this disorder is observed after lesions of the superior temporal cortex¹, a puzzling discrepancy between the species. Here we show that, contrary to the widely accepted view, the superior temporal cortex is the neural substrate of spatial neglect in humans, as it is in monkeys. Unlike the monkey brain, spatial awareness in humans is a function largely confined to the right superior temporal cortex, a location topographically reminiscent of that for language on the left². Hence, the decisive phylogenetic transition from monkey to human brain seems to be a restriction of a formerly bilateral function to the right side, rather than a shift from the temporal to the parietal lobe. One may speculate that this lateralization of spatial awareness parallels the emergence of an elaborate representation for language on the left side.

Spatial neglect is a characteristic failure to explore the side of space contralateral to a brain lesion. Patients with this disorder behave as if one side of the surrounding space had ceased to exist. Since the early post-mortem studies, we have believed that, in humans, lesions located predominantly in the posterior parietal lobe are critical for this disorder. Analyses of computerized tomography scans of right-hemispheric stroke patients with neglect found that superimposed lateral projections of these lesions centred on the inferior parietal lobule (IPL)^{3,4} and the temporo-parieto-occipital (TPO) junction⁴. More recent studies have confirmed the validity of this conclusion although evidence for additional pathology leading

to spatial neglect outside this region has also been reported⁵⁻⁸. However, in monkeys, evidence that spatial neglect is associated with such lesions has been less convincing. Ablation studies revealed no or only minimal spatial neglect⁹⁻¹³. These observations raised a controversial discussion on the anatomical and functional homology between human and monkey parietal cortex. In view of this seeming discrepancy, we re-assessed the evidence for involvement of the inferior parietal cortex and TPO junction in human spatial neglect.

Our approach was notably different from previous studies of neglect patients because we focused on patients suffering from 'pure' spatial neglect, that is on patients not having additional primary defects in their visual field. This allowed us to study the neural correlate of pure spatial neglect without a bias to posterior visual regions inducing these latter defects. Forty-nine acute stroke patients with severe spatial neglect but no visual-field defects following circumscribed unilateral right-hemispheric brain lesions, consecutively admitted over a five-year period from a well defined recruitment area, were investigated. In 33 of these patients we found lesions affecting cortical areas, whereas lesions in the other 16 patients were confined to subcortical structures, namely the basal ganglia or the thalamus. Of the patients with cortical lesions, 25 (Table 1) showed no concurrent involvement of the basal ganglia or the thalamus; that is, they showed no involvement of those subcortical nuclei known to induce neglect independent of concomitant cortical damage^{3,7,14,15}. The superimposed lesion plots of these 25 subjects were compared with 25 right-brain-damaged (RBD) controls (Fig. 1) who likewise had no visual-field defects and cortical lesions without involvement of the basal ganglia or the thalamus. In clear contrast to controls, the lesion overlap in the neglect patients centred on the superior temporal gyrus (STG; Brodmann areas 22 and 42). Neighbouring structures affected were the ventral postcentral gyrus and the operculum. Interestingly, we found no evidence for a predominant involvement of the IPL^{3,4,6,7}, the TPO junction^{4,7,8}, the cingulate cortex^{7,8} or the middle temporal gyrus⁵ in patients with pure spatial neglect. This pattern did not change when we included the eight patients suffering from cortical lesions but with concurrent involvement of the basal ganglia or the thalamus.

For statistical comparison of IPL with STG involvement, we defined the IPL and STG areas in Talairach space¹⁶ and determined (using MRIcro software¹⁷) to which percentage the individual lesion affected these two regions of interest (Fig. 1d). Between-group comparisons revealed a highly significant difference for the STG (Mann-Whitney $U = 105$, $P < 0.001$) showing that the area of the STG involved was 6.9 times larger in neglect patients (Fig. 1d). There was no such significant difference for the IPL ($U = 301$, $P = 0.801$). The comparison of relative involvement of the STG and the IPL within each group revealed a 3.7 times larger involvement of the STG in neglect patients (Wilcoxon $Z = -3.51$, $P < 0.001$). In controls, this comparison was not significantly different ($Z = -1.07$, $P = 0.285$). Table 2 in Supplementary Information gives coefficients for the correlation of IPL and STG involvement and the behavioural performance of the patients.

The previous results on lesion location in neglect may have resulted from the inclusion of a significant proportion of patients who suffered not only from spatial neglect but also from additional visual-field defects. Hemianopia was present, for example, in 87% of the patients with spatial neglect studied by Vallar and Perani³ and in 50% of the patients investigated by Perenin⁶. Hence, it is plausible that in many cases lesions involved posterior visual regions, possibly confounding the interpretation of lesion location relevant to spatial neglect. To test this assumption, we studied 11 acute unilateral right-hemispheric stroke patients suffering from both severe spatial neglect and additional visual-field defects who were consecutively admitted over a 1.5-year period from the same recruitment area as for the groups above. Three of these patients showed concurrent

involvement of the basal ganglia or thalamus and were not considered in the present analysis (Table 1). During the same period, we found four patients who suffered from an acute right-hemispheric stroke in the same vascular region as the neglect patients with visual-field defects and likewise exhibited visual-field defects but no spatial neglect (Table 1).

If previous conclusions on lesion location in neglect were confounded by concomitant damage of visual cortical structures, the present comparison should reveal the following: (1) a clear difference in lesion location between those patients with pure neglect and those with both neglect and hemianopia (the latter involving also posterior regions); (2) those areas found as the neural correlate of spatial neglect in patients with pure neglect (Fig. 1) are affected in the patients suffering from both neglect and hemianopia; (3) the centre of lesion overlap in patients with neglect and hemianopia is comparable to that reported in earlier studies (that is, involves the IPL and TPO junction); and (4) the posterior parts of lesion overlap in the patients with neglect and hemianopia also overlap in those patients without neglect but with hemianopia due to a stroke in the territory of the same cerebral artery. Our data were in full accor-

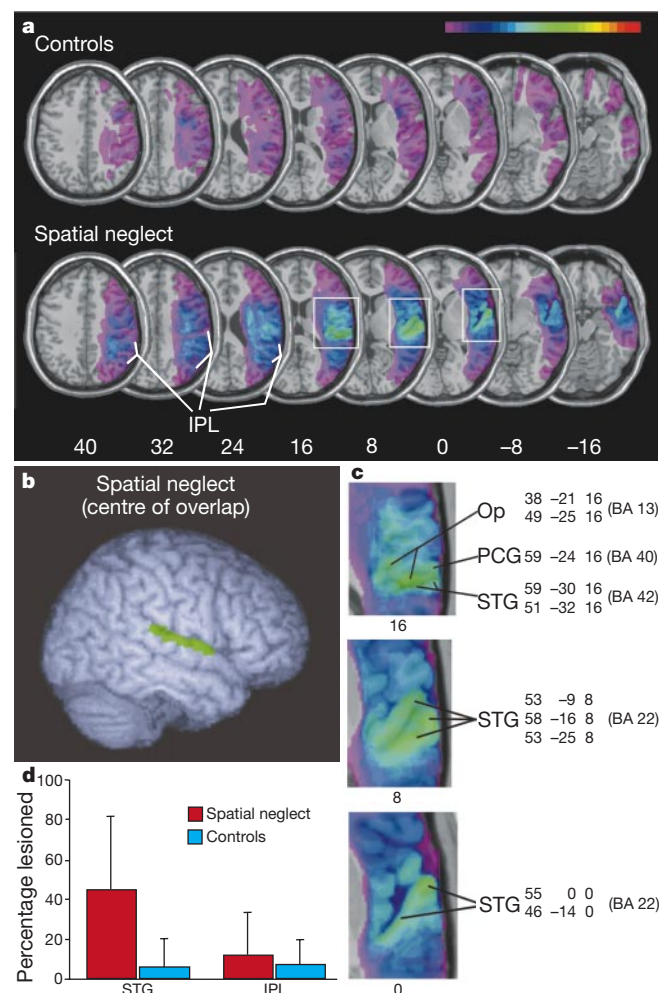


Figure 1 Lesion analysis of patients without (controls) and with spatial neglect. **a**, Overlay plots. Talairach z-coordinates¹⁶ of the transverse sections are given. The number of overlapping lesions is illustrated by colour, from violet ($n = 1$) to red ($n = 25$). In line with the criterion used previously³, the centre of overlap was defined as those voxels that showed lesions in more than 15 patients (green area). IPL, inferior parietal lobule.

b, c, Surface view (**b**) and exploded view (**c**) of the centre of lesion overlap. Talairach coordinates¹⁶ of the locations marked are given. Op, operculum; PCG, ventral postcentral gyrus; STG, superior temporal gyrus; BA, Brodmann area. **d**, Mean percentage of area showing lesions within the STG and within the IPL.

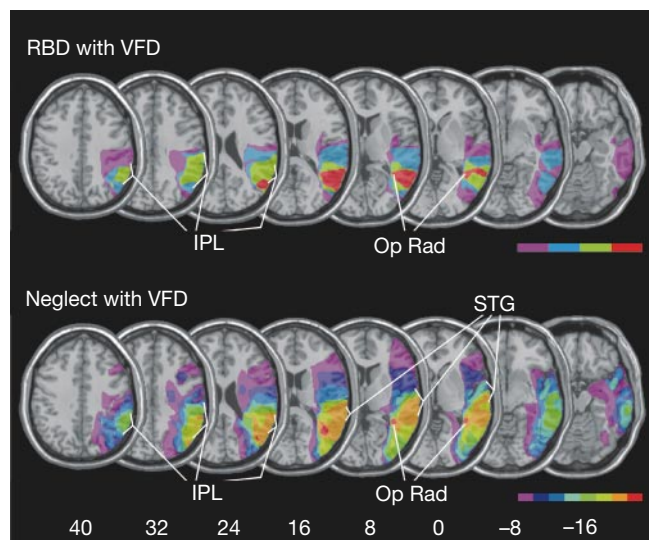


Figure 2 Lesion analysis of patients with visual-field defects (VFDs), without (RBD) and with spatial neglect. Talairach z-coordinates¹⁶ of the transverse sections are given. The number of overlapping lesions is illustrated by colour, from violet to red ($n = 1-4$ for RBD, $n = 1-8$ for neglect). The centre of overlap was defined as those voxels that showed lesions in more than 50% of the subjects (green–red area for RBD; green–orange–red area for neglect). IPL, inferior parietal lobule; Op Rad, optic radiation; STG, superior temporal gyrus.

dance with these expectations. The centre of lesion overlap in the patients with both spatial neglect and visual-field defects involved the IPL and the TPO junction, which corresponds to earlier reports^{3,4,6-8} (Fig. 2, lower panel). However, the same region was also affected in the patients without neglect but with visual-field defects following a stroke in the territory of the middle cerebral artery (Fig. 2, upper panel). In both groups, visual-field defects were caused by lesion extension to the posterior horn of the lateral ventricle, thereby affecting the optic radiation.

The present finding fits with observations of the consequences of experimental lesions in monkeys. Watson and co-workers¹ found spatial neglect after ablation around both banks of the superior temporal sulcus. Although the focus found here in humans was on the superior temporal gyrus (STG), the observations correspond in

that the superior temporal cortex rather than the IPL or the TPO junction is the substrate of spatial neglect in both monkeys and humans. The STG is located at the transition between the two major pathways of cortical visual processing, the 'what' and 'where' systems, respectively¹⁸. The STG is known to receive polysensory input from both streams thus representing a site of multimodal sensory convergence¹⁹⁻²². Our findings indicate that this information may serve as a matrix for spatial exploration and for spatial awareness. The results further indicate that the cytoarchitecturally identical areas 7 in the monkey and in the human are also functionally homologous. In monkeys, lesion of this area in either the right or the left hemisphere causes misreaching for objects with the contralesional arm but not spatial neglect^{10,11}. The same is true for area-7 lesions in humans⁶. However, unlike area 7 in the parietal cortex, left- and right-hemisphere functions of the STG are different in monkeys and humans. Whereas loss of awareness of the contralesional side in monkeys is observed after lesions of this area in the left or right hemisphere^{1,23}, spatial neglect in humans is predominantly associated with right-hemispheric lesions²⁴. The human superior temporal cortex in the opposite, left hemisphere subserves language functions². Hence, the phylogenetic transition from the monkey to the human brain seems to be a restriction of a formerly bilateral function to the right side, rather than a shift from the temporal to the posterior parietal lobe. One may speculate that this lateralization of spatial awareness parallels the emergence of an elaborate representation of language on the left. However, in contrast to previous conviction, we suggest that parietal lobe organization (probably including human areas 39 and 40 (ref. 25)) is in principle comparable and functionally homologous in both species, directly coding space for action. □

Methods

Clinical investigation

Patients were classified as having spatial neglect when they showed the typical clinical behaviour such as spontaneous deviation of the head and eyes toward the ipsilesional side, orienting towards the ipsilesional side when addressed from the front or the left, and ignoring contralesionally located people or objects. In addition, all patients were further assessed with the following four clinical neglect tests and had to fulfil the criterion in at least two of them.

Letter cancellation test²⁶: sixty target letters were distributed amid distractors on a horizontally oriented 21 × 29.7 cm sheet of paper, 30 within each half field. Patients had to cancel all target letters and were classified as suffering from spatial neglect when they omitted at least five left-sided targets. Bells test²⁷: this consists of seven columns each containing five targets (bells) among distractors. Three of the seven columns (15 targets) are on the left side of a horizontally oriented 21 × 29.7 cm sheet of paper. More than five

Table 1 Demographic and clinical data of patients with spatial neglect and without (controls, RBD)

			Spatial neglect	Controls	Spatial neglect	RBD
			No visual-field defects		With visual-field defects	
Number			25	25	8	4
Sex			14 f, 11 m	11 f, 14 m	5 f, 3 m	1 f, 3 m
Age (yr)		Median (range)	68 (31–89)	63 (33–77)	71 (55–83)	63 (60–79)
Aetiology			20 infarct	21 infarct	7 infarct	3 infarct
			5 haemorrhage	4 haemorrhage	1 haemorrhage	1 haemorrhage
Time since lesion (d)		Median (range)	9 (1–140)	3.5 (1–54)	8 (3–157)	7 (3–22)
Paresis of contralesional side*			100	84	88	50
% present	Arm	Median (range)	2 (0–4.5)	3.5 (0–5)	2.5 (0–5)	3.3 (1.5–5)
	Leg	Median (range)	4 (0–5)	5 (2–5)	4.5 (0–5)	4.8 (4.5–5)
Somatosensory deficit of contralesional side (touch) (% present)	Arm		64	36	88	50
	Leg		64	28	75	50
Visual-field defect (% present)			0	0	100	100
					(7 Ha, 1 Qa)	(2 Ha, 2 Qa)
Letter cancellation	Left	Median (range)	1 (0–29)	29 (16–30)	0 (0–21)	29.5 (29–30)
	Right	Median (range)	22 (4–30)	30 (17–30)	18 (11–26)	30 (28–30)
Bells test	Left	Median (range)	0 (0–14)	14 (8–15)	0 (0–11)	15 (14–15)
	Right	Median (range)	11 (2–15)	15 (7–15)	7 (4–14)	15 (14–15)
Baking tray task	Left	Median (range)	4.5 (0–9)	8 (6–10)	2.5 (0–6)	8†
	Right	Median (range)	11 (7–16)	8 (6–10)	13.5 (10–16)	8†
Copying (% omitted)		Median (range)	50 (0–88)	0 (0–25)	75 (38–88)	0†

RBD, right brain damage; f, female; m, male; Ha, hemianopia; Qa, quadrantanopia.

* Paresis was scored with the usual clinical ordinal scale: 0, no trace of movement; to 5, normal movement.

† No variation in data.

left-sided omissions indicated spatial neglect. Baking tray task²⁸: patients had to place 16 identical items as evenly as possible on a blank test sheet (21 × 29.7 cm). Any distribution that is more skewed than seven items in the left half and nine on the right²⁸ was considered a sign of neglect. Copying task: patients were asked to copy a complex multi-object scene consisting of four figures on a 21 × 29.7 cm sheet of paper. Omission of at least one of the left-sided features of each figure was scored as one and omission of each whole figure was scored as two. One additional point was given when left-sided figures were drawn on the right side. The maximum score was eight. A score higher than one, that is, more than 12.5% omissions, indicated neglect.

All other relevant demographic and clinical parameters are shown in Table 1, together with an overview of these data. Visual-field defects were measured by Tübingen perimetry and standardized neurological examination.

Lesion analysis

Brain lesions were identified by computerized tomography or magnetic resonance imaging (MRI). Patients with diffuse or bilateral brain lesions, patients with tumours and patients in whom imaging revealed no manifest lesion were excluded. Lesions were mapped with MRIcro software¹⁷ (<http://www.psychology.nottingham.ac.uk/staff/cr1/mricro.html>). They were drawn manually on slices of a template MRI scan from the Montreal Neurological Institute (http://www.bic.mni.mcgill.ca/cgi/icbm_view), which is based on 27 T1-weighted MRI scans, normalized to Talairach space¹⁶. This scan was distributed with SPM99 (<http://www.fil.ion.ucl.ac.uk/spm/spm99.html>). For superimposing of the individual brain lesions, the same MRIcro software¹⁷ was used. Three-dimensional rendering was carried out with mri3dX software (<http://mrrc11.mrrc.liv.ac.uk/mri3dX>).

Received 23 January; accepted 25 April 2001.

- Watson, R. T., Valenstein, E., Day, A. & Heilman, K. M. Posterior neocortical systems subserving awareness and neglect. *Arch. Neurol.* **51**, 1014–1021 (1994).
- Binder, J. The new neuroanatomy of speech perception. *Brain* **123**, 2371–2372 (2000).
- Vallar, G. & Perani, D. The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia* **24**, 609–622 (1986).
- Heilman, K. M., Watson, R. T., Valenstein, E. & Damasio, A. R. in *Localization in Neuropsychology* (ed. Kertesz, A.) 471–492 (Academic, New York, 1983).
- Samuelsson, H., Jensen, C., Ekholm, S., Naver, H. & Blomstrand, C. Anatomical and neurological correlates of acute and chronic visuospatial neglect following right hemisphere stroke. *Cortex* **33**, 271–285 (1997).
- Perenin, M. T. in *Parietal Lobe Contributions to Orientation in 3D Space* (eds Thier, P. & Karnath, H.-O.) 289–308 (Springer, Heidelberg, 1997).
- Leibovitch, F. S. et al. Brain-behavior correlations in hemispatial neglect using CT and SPECT: the Sunnybrook Stroke Study. *Neurology* **50**, 901–908 (1998).
- Leibovitch, F. S. et al. Brain SPECT imaging and left hemispatial neglect covaried using partial least squares: the Sunnybrook Stroke Study. *Hum. Brain Mapp.* **7**, 244–253 (1999).
- Eitlinger, G. & Kalsbeek, J. E. Changes in tactile discrimination and in visual reaching after successive and simultaneous bilateral posterior parietal ablations in the monkey. *J. Neurol. Neurosurg. Psychiatry* **25**, 256–268 (1962).
- Lamotte, R. H. & Acuna, C. Deficits in accuracy of reaching after removal of posterior parietal cortex in monkeys. *Brain Res.* **139**, 309–326 (1978).
- Faugier-Grimaud, S., Frenois, C. & Stein, D. G. Effects of posterior parietal lesions on visually guided behavior in monkeys. *Neuropsychologia* **16**, 151–168 (1978).
- Lynch, J. C. & McLaren, J. W. Deficits of visual attention and saccadic eye movements after lesions of parietooccipital cortex in monkeys. *J. Neurophysiol.* **61**, 74–90 (1989).
- Gaffan, D. & Hornak, J. Visual neglect in the monkey. Representation and disconnection. *Brain* **120**, 1647–1657 (1997).
- Damasio, A. R., Damasio, H. & Chui, H. C. Neglect following damage to frontal lobe or basal ganglia. *Neuropsychologia* **18**, 123–132 (1980).
- Motomura, N. et al. Unilateral spatial neglect due to hemorrhage in the thalamic region. *Acta Neurol. Scand.* **74**, 190–194 (1986).
- Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System—An Approach to Cerebral Imaging*. (Thieme, New York, 1988).
- Rorden, C. & Brett, M. Stereotaxic display of brain lesions. *Behav. Neurol.* (in press).
- Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, Massachusetts, 1982).
- Jones, E. G. & Powell, T. P. S. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* **93**, 793–820 (1970).
- Seltzer, B. & Pandya, D. N. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* **149**, 1–24 (1978).
- Bruce, C., Desimone, R. & Gross, C. G. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384 (1981).
- Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
- Luh, K. E., Butter, C. M. & Bachtel, H. A. Impairments in orienting to visual stimuli in monkeys following unilateral lesions of the superior sulcal polysensory cortex. *Neuropsychologia* **24**, 461–470 (1986).
- Mesulam, M.-M. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Phil. Trans. R. Soc. Lond. B* **354**, 1325–1346 (1999).
- Darling, W. G., Rizzo, M. & Butler, A. J. Disordered sensorimotor transformations for reaching following posterior cortical lesions. *Neuropsychologia* **39**, 237–254 (2001).
- Weintraub, S. & Mesulam, M.-M. in *Principles of Behavioral Neurology* (ed. Mesulam, M.-M.) 71–123 (Davis, Philadelphia, 1985).
- Gauthier, L., Dehaut, F. & Joannette, Y. The bells test: a quantitative and qualitative test for visual neglect. *Int. J. Clin. Neuropsychol.* **11**, 49–54 (1989).

- Tham, K. & Tegnér, R. The baking tray task: a test of spatial neglect. *Neuropsychol. Rehab.* **6**, 19–25 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie awarded to H.-O.K. We thank M. Niemeier, L. Johannsen and U. Zimmer for support with the neuropsychological testing of the patients; P. Thier for discussion and suggestions for the manuscript; U. Amann for help in the tomography archives; and C. Rorden for developing the MRIcro software.

Correspondence and requests for materials should be addressed to H.-O.K. (e-mail: karnath@uni-tuebingen.de).

Single neurons in prefrontal cortex encode abstract rules

Jonathan D. Wallis, Kathleen C. Anderson & Earl K. Miller

Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The ability to abstract principles or rules from direct experience allows behaviour to extend beyond specific circumstances to general situations. For example, we learn the 'rules' for restaurant dining from specific experiences and can then apply them in new restaurants. The use of such rules is thought to depend on the prefrontal cortex (PFC) because its damage often results in difficulty in following rules¹. Here we explore its neural basis by recording from single neurons in the PFC of monkeys trained to use two abstract rules. They were required to indicate whether two successively presented pictures were the same or different depending on which rule was currently in effect. The monkeys performed this task with new pictures, thus showing that they had learned two general principles that could be applied to stimuli that they had not yet experienced. The most prevalent neuronal activity observed in the PFC reflected the coding of these abstract rules.

Neurons in the prefrontal cortex (PFC) encode many different types of information from all stages of the perception–action cycle². They are activated by stimuli from all sensory modalities^{3–5}, before and during a variety of actions^{6–8}, during memory for past events⁹, in anticipation of expected events and behavioural consequences^{10–12}, and are modulated by 'internal' factors such as motivational and attentional state^{13,14}. The PFC is thought to use this diverse information for the 'higher order' control of behaviour, in particular the application of behaviour-guiding rules that are lost after damage to the PFC^{15,16}. Although rules can be specific and concrete (for example, 'red' means 'stop'), it is the abstraction of general rules or principles (those not tied to any particular stimulus or behavioural response) that allows for the flexibility and adaptability that are central to intelligent behaviour. Although recent studies indicate that PFC neurons can encode concrete rules between specific stimuli and behavioural responses^{17–19}, we do not know how, or even whether, PFC neurons can encode abstract rules.

Thus, we trained two monkeys to switch flexibly between two abstract rules. The 'match' rule required monkeys to release a lever if two successively presented (sample and test) objects were identical, whereas the 'nonmatch' rule required the lever release if the two objects were different (Fig. 1). The rule applicable for each trial was randomly indicated by a cue that was presented with the sample. To separate the neural activity related to the physical properties of the

cue from the rule that it signified, two distinct cues from different sensory modalities were used to indicate the same rules, whereas cues signifying different rules were from the same modality (Fig. 1). Both monkeys were proficient at the task (92% and 84% correct

performance) and performed well above chance when applying the rules the very first time they encountered a new object (70% correct, 4 objects $\times$ 55 recording sessions = 220 objects; $P < 10^{-8}$; binomial test).

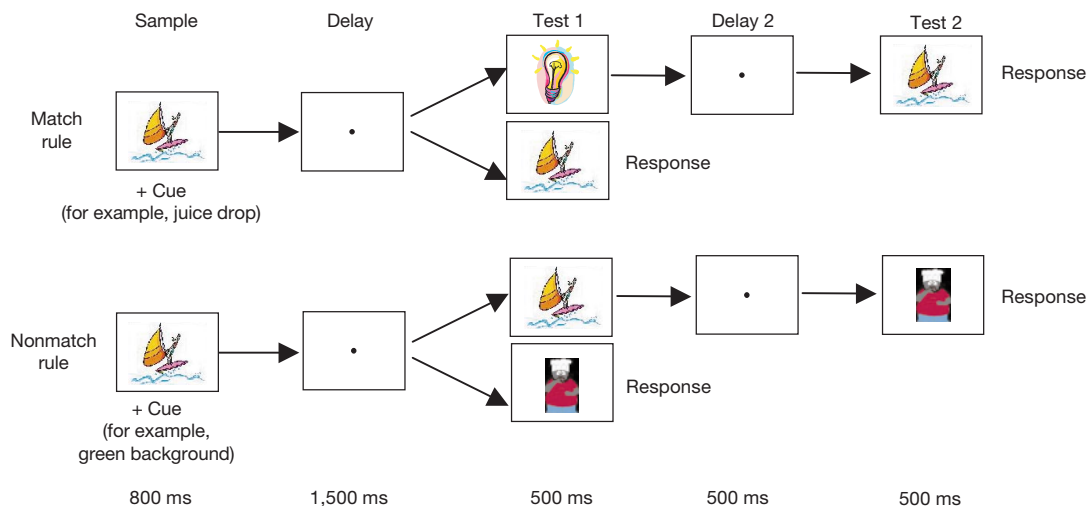


Figure 1 The behavioural task. Monkeys grasped a lever and maintained central fixation. A sample object was followed by a brief delay, and then by only one test object. Illustrated are two trial types for each rule (bifurcating arrows). For the match rule, the monkeys released a lever if the test object matched the sample. For the nonmatch rule, they released the lever if the test object did not match. Otherwise, they held the lever through a

second delay until appearance of a second test object that always required a response. Thus, only the first test required a decision; the second delay and test was used so that a behavioural response was required on each trial, ensuring that monkeys were always paying attention.

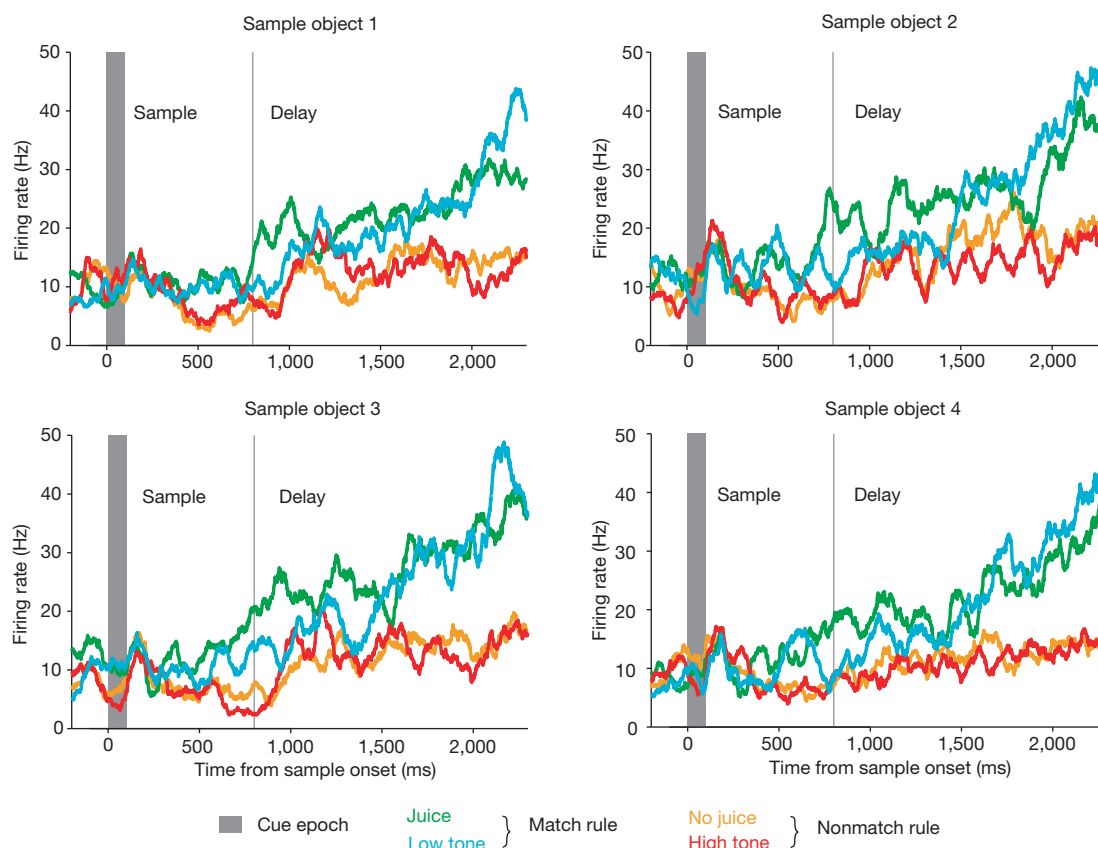


Figure 2 A neuron exhibiting rule selectivity. The neuron shows greater activity during match trials, regardless of which cue signified the rule or which object was remembered. It showed a 72% difference in activity between the rules in the sample epoch and a 120% difference during the delay (0.27 and 0.37 as measured by the selectivity index). The

mean firing rate during the delay epoch was 13.2 spikes s^{-1} and the 99% confidence interval was ± 0.91 for the nonmatch rule and 24.8 spikes s^{-1} (± 1.15) for the match rule.

We recorded the activity of 492 neurons from the dorsolateral, ventrolateral and orbitofrontal PFC. To discern whether neural activity in the sample and delay epochs reflected the cues, sample objects, and/or rules, three-way analysis of variance (ANOVA) tests were computed for each neuron (see Methods). The modal group of PFC neurons showed activity that reflected the current rule (200/492 or 41%; see Table 1). Figure 2 shows a good example of a rule-selective neuron that exhibited greater activity when the match rule was cued. This activity cannot be explained by the physical characteristics of the sample object or cues; activity was equivalent regardless of the specific sample or of which cue signified a given rule. It cannot be related to anticipation of the behavioural response as the monkey could not know if the forthcoming object would require a response. Nor could it be related to differences in reward expectation. Although in one of the conditions the rule was cued with a drop of juice at the beginning of the sample epoch, the expectation of reward was identical for all conditions for the remainder of the trial. Furthermore, performance on match and nonmatch trials was virtually identical (across monkeys, average error rates differed by only 0.1% and reaction times by 7 ms). Thus, the most parsimonious explanation is that the activity reflected the abstract rules that the monkeys were using to guide their behaviour.

Figure 3 shows a distribution of the magnitude of rule-selectivity for the population of rule-selective neurons (see Methods). The mean of their absolute values was 0.18 and 0.13 for the sample and delay, respectively, which corresponds to a 48% and 33% difference in activity for match versus nonmatch rules. The number of neurons that showed stronger activity for the match rule (101/200, or 50.5%) was similar to that of neurons that were more active for the nonmatch rule (99/200, or 49.5%) and the magnitude of rule-selectivity to each did not differ (Wilcoxon rank sum test, $P > 0.05$ in both sample and delay intervals).

The second most prevalent type of neuronal activity observed was a Cue $\times$ Rule interaction (167/492 or 34% of neurons), which occurred when a neuron was most active to a single cue. This may simply reflect the physical properties of the cue, although, in principle, it could also carry some rule information—for example, by encoding rule information but only from a single modality. Finally, 14% of the neurons (70/492) showed activity that reflected the identity of the sample object.

To further demonstrate abstract rule representation, pairwise

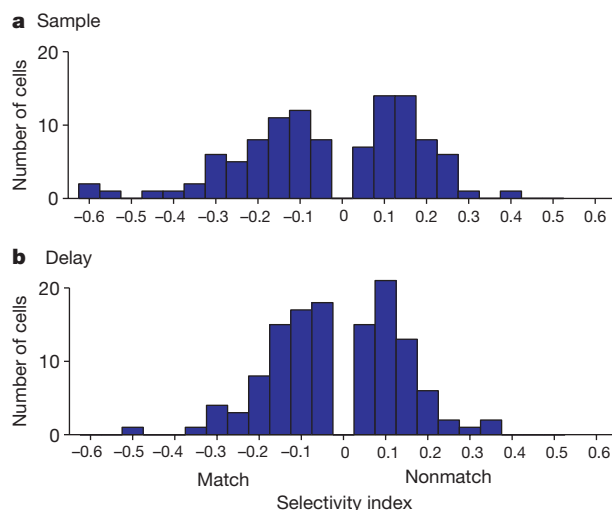


Figure 3 Distribution of the magnitude of rule effect. **a, b**, Data from neurons that were rule-selective during the sample epoch (107 neurons; **a**) and delay (127 neurons; **b**). Match and nonmatch refer to the values for neurons exhibiting stronger activity to the match or nonmatch rule.

t -tests were made for each neuron between activity to the four cues (yielding six unique comparisons) across both the sample and delay epochs ($P < 0.01$, Bonferroni corrected). Because our task design set cue modality in opposition to rule (two cues from the same sensory modality instructed two different rules and the same rule was indicated by cues from two different modalities) many PFC neurons tended to group cues by modality or rule (Table 1). We adopted a strict criterion to indicate such selectivity: for a given set of cue pairs, there had to be significant differences in all across-pair comparisons (4/6 comparisons) but no significant difference to cues within each of the pairs (2/6 comparisons). Of the 91 PFC neurons that met this criterion, most (69/91 or 76%) grouped the cues by rules. Their activity was significantly different to similar cues that indicated different rules but not significantly different to distinct cues that indicated the same rule. Far fewer neurons grouped the cues on the basis of their modalities (15/91, or 16%) or by neither rule nor modality (6/91 or 6%). Further evidence of supramodal rule-coding is the similar proportion of rule-selective neurons in the two monkeys even though different cue modalities were used for each monkey (125/303 or 41% in monkey A, and 75/189 or 40% in monkey B).

Encoding of abstract rules was evident throughout the PFC (Fig. 4). During the sample epoch there was a higher incidence of

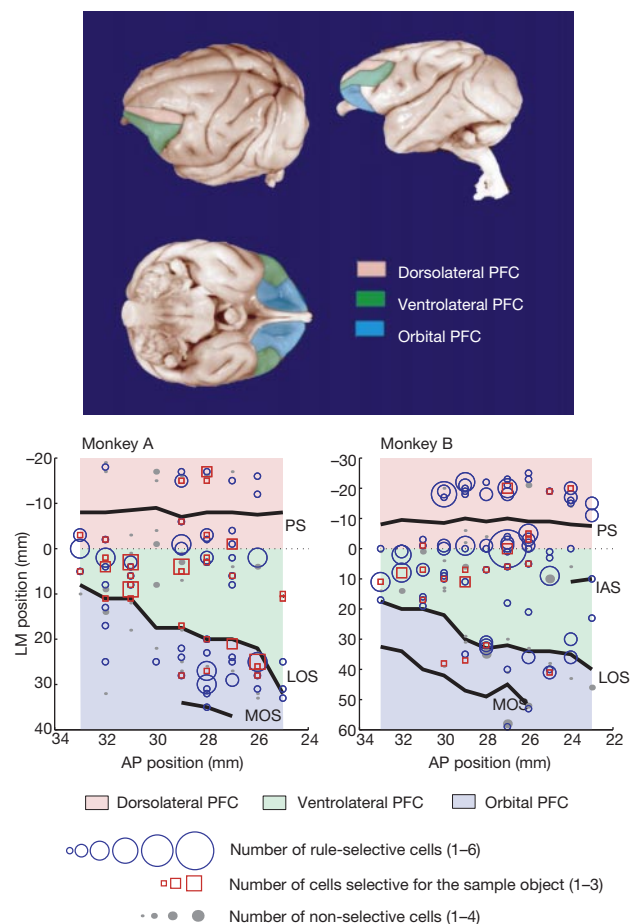


Figure 4 Location of neurons showing rule or object selectivity in either the delay or sample. The anterior–posterior axis was defined with respect to the interaural line, whereas the lateral–medial axis was defined with respect to the ventral lip of the principal sulcus (shown with the dotted line). Dorsolateral PFC (pink) was the cortex lying dorsal to the ventral lip of the principal sulcus. Ventrolateral PFC (green) was between the ventral lip of the principal sulcus and the depth of the lateral orbital sulcus. Orbitofrontal cortex (blue) was medial to the depth of the lateral orbital sulcus. PS, principal sulcus; IAS, inferior arcuate sulcus; LOS, lateral orbital sulcus; MOS, medial orbital sulcus.

Table 1 Neuronal selectivity in different task periods

	Sample epoch				Delay epoch				Either epoch			
	D	V	O	Total	D	V	O	Total	D	V	O	Total
% of cells selective for												
Cue	19	19	17	19	15	8	6	11	31	24	23	27
Rule	29	16	18	22	29	24	23	26	49	37	32	41
Sample object	7	18	8	11	4	12	4	7	10	23	10	14
Cue × Rule	24	31	21	25	14	17	11	14	33	40	28	34
Rule × Sample object	0	3	0	1	1	1	0	1	1	4	0	2
Cue × Sample object	2	0	1	1	3	1	2	2	4	1	2	2

D, dorsolateral PFC ($n = 197$); V, ventrolateral PFC ($n = 169$); O, orbitofrontal PFC ($n = 126$); total, all three areas combined ($n = 492$).

rule-selective neurons in the dorsolateral than in either the ventrolateral or orbital PFC (57/197 or 29%, 27/169 or 16%, 23/126 or 18%, respectively; see Table 1, $\chi^2 = 9.25$, $P < 0.01$) but by the delay, rule-selectivity was equally prevalent (58/197 or 29%, 40/169 or 24%, 29/126 or 23%, respectively; $\chi^2 = 1.85$, $P > 0.1$). The magnitude of rule-selectivity did not differ between the PFC regions (Kruskal–Wallis, $P > 0.1$). There was no difference in the distribution of neurons preferring the match or the nonmatch rules ($\chi^2 = 0.3$, $P > 0.1$); they were evenly split in each PFC region. In fact, neurons preferring different rules were often recorded from the same electrode. Sample-selective neurons were evident in all three regions, but were more numerous in the ventrolateral PFC during both the sample epoch (ventrolateral, 30/169 or 18%; dorsolateral, 13/197 or 7%; orbital 10/126 or 8%; $\chi^2 = 11.8$, $P < 0.01$) and the delay (20/169 or 12%, 7/197 or 4%, 5/126 or 4% respectively, $\chi^2 = 10.3$, $P < 0.01$).

The capacity for abstraction is an important component of higher cognition; it frees an organism from specific associations and gives it the ability to generalize and develop overarching concepts and principles. The ability of PFC neurons to group cues into behavioural categories that are dependent on abstract rules is consistent with observations of a loss of flexibility after PFC damage and with the ability of PFC neurons to form perceptual categories²⁰. The prevalence of rule activity is not inconsistent with studies showing the role of the lateral PFC in working memory^{21–24} or the orbital PFC in processing affective information^{12,25,26}, but it does suggest that the abstraction of rules and principles may be an important prefrontal function. □

Methods

Behavioural and recording methods

Trials were randomized and balanced across all relevant features (cues, samples, rules and so on). Monkeys completed about 1,000 trials per day at a consistent level of performance. Eye position remained within 1.7 degrees of the fixation spot throughout the trial and was monitored with an infrared system (ISCAN). Breaks in fixation were not counted in the error rates. The pattern of microsaccades (small eye movements) was similar for different rules. Recordings were made from the PFC of two adult rhesus monkeys (*Macaca mulatta*) using arrays of eight tungsten microelectrodes (FHC Instruments) using a grid (Crist Instruments) with 1-mm spacing. Recordings were localized using magnetic resonance imaging and neurons were randomly sampled; no attempt was made to select neurons on the basis of responsiveness. neural waveforms were digitized and analysed offline using principal components (Plexon Systems).

Four new objects were chosen each day and used throughout a recording session. Use of four samples meant that the identity of the nonmatching test object could not be predicted and thus monkeys needed to remember the current sample and rule. The rule was signified by a brief (100 ms) cue coincident with sample onset (for monkey A, a drop of juice or a blue background indicated the match rule, whereas no juice or a green background indicated the nonmatch rule; for monkey B, juice or a low tone indicated the match rule, whereas no juice or a high tone indicated the nonmatch rule). The sound of the juice delivery solenoid was masked by white noise. Both monkeys performed somewhat better (about 9%) for the juice cues but reaction times did not differ among cues. There was no bias toward responding to the first or second test object (error rates were similar), but monkeys responded about an average of 35 ms faster to the second test, presumably because the response was predictable.

Data analysis

Only data from correct trials were used. Sample activity was summed from 200 ms after sample onset to its offset 600 ms later. Delay activity was summed over the entire delay epoch. All analyses used data from only the sample and first delay epochs although neural

activity during the second delay was similar to that seen during the first. The three-way ANOVA was a standard, non-nested, linear model with two levels of interactions and was evaluated at $P < 0.01$. Factors included cue modality (juice versus background colour for monkey A; juice versus tones for monkey B), rule (match versus nonmatch) and the sample object. Rule-selective neurons showed a main effect of rule and no interaction with cue or sample. Likewise, cue or sample-selective neurons showed main effects and no interactions. Magnitude of selectivity was calculated using a standard index (activity to nonmatch minus match rule divided by their sum) and converted to a percentage difference.

Received 20 February; accepted 9 May 2001.

- Milner, B. Effects of different brain lesions on card sorting. *Arch. Neurol.* **9**, 100–110 (1963).
- Fuster, J. M. *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe* (Lippincott, Williams & Wilkins, Philadelphia, 1997).
- Watanabe, M. Reward expectancy in primate prefrontal neurons. *Nature* **382**, 629–632 (1996).
- Rao, S. C., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824 (1997).
- Fuster, J. M., Bodner, M. & Kroger, J. K. Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* **405**, 347–351 (2000).
- Fuster, J. M. Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *J. Neurophysiol.* **36**, 61–78 (1973).
- di Pellegrino, G. & Wise, S. P. Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J. Neurosci.* **13**, 1227–1243 (1993).
- Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407 (1998).
- Goldman-Rakic, P. C. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
- Leon, M. I. & Shadlen, M. N. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* **24**, 415–425 (1999).
- Rainer, G., Rao, S. C. & Miller, E. K. Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.* **19**, 5493–5505 (1999).
- Tremblay, L. & Shultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
- Rolls, E. T., Xaxley, S. & Sienkiewicz, Z. J. Gustatory responses of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *J. Neurophysiol.* **64**, 1055–1066 (1990).
- Rainer, G., Asaad, W. F. & Miller, E. K. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* **393**, 577–579 (1998).
- Wise, S. P., Murray, E. A. & Gerfen, C. R. The frontal cortex–basal ganglia system in primates. *Crit. Rev. Neurobiol.* **10**, 317–356 (1996).
- Miller, E. K. The prefrontal cortex: complex neural properties for complex behavior. *Neuron* **22**, 15–17 (1999).
- Hoshi, E., Shima, K. & Tanji, J. Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *J. Neurophysiol.* **80**, 3392–3397 (1998).
- White, I. M. & Wise, S. P. Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.* **126**, 315–335 (1999).
- Asaad, W. F., Rainer, G. & Miller, E. K. Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.* **84**, 451–459 (2000).
- Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
- Fuster, J. M. & Alexander, G. E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
- Niki, H. Prefrontal unit activity during delayed alternation in the monkey. II. Relation to absolute versus relative direction of response. *Brain Res.* **68**, 197–204 (1974).
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
- Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
- Rolls, E. T. The orbitofrontal cortex and reward. *Cereb. Cortex* **10**, 284–294 (2000).
- Roberts, A. C. & Wallis, J. D. Inhibitory control and affective processing in the prefrontal cortex: neuropsychological studies in the common marmoset. *Cereb. Cortex* **10**, 252–262 (2000).

Acknowledgements

We thank W. Asaad, D. Freedman, C. Kiddoo, M. Warden and M. Wicherski for valuable comments. This work was supported by a NINDS grant, a NIMH Conti Center grant, the RIKEN-MIT Neuroscience Research Center and the Class of 1956 Chair (E.K.M.). J.D.W. was supported by the Wellcome Trust.

Correspondence and requests for materials should be addressed to E.K.M. (e-mail: ekm@ai.mit.edu).

Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P₂-mediated inhibition

Huai-hu Chuang^{*†}, Elizabeth D. Prescott^{*†}, Haeyoung Kong[‡], Shannon Shields[§], Sven-Eric Jordt^{*}, Allan I. Basbaum[§], Moses V. Chao[‡] & David Julius^{*}

^{*} Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143-0450, USA

[‡] Molecular Neurobiology Program, Skirball Institute of Biomolecular Medicine, New York University School of Medicine, New York, New York 10016, USA

[§] Departments of Anatomy and Physiology and the W. M. Keck Center for Neuroscience, University of California, San Francisco, California 94143-0444, USA

[†] These authors contributed equally to this work.

Tissue injury generates endogenous factors that heighten our sense of pain by increasing the response of sensory nerve endings to noxious stimuli^{1,2}. Bradykinin and nerve growth factor (NGF) are two such pro-algesic agents that activate G-protein-coupled (BK₂) and tyrosine kinase (TrkA) receptors, respectively, to stimulate phospholipase C (PLC) signalling pathways in primary afferent neurons^{3,4}. How these actions produce sensitization to physical or chemical stimuli has not been elucidated at the molecular level. Here, we show that bradykinin- or NGF-mediated potentiation of thermal sensitivity *in vivo* requires expression of VR1, a heat-activated ion channel on sensory neurons. Diminution of plasma membrane phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) levels through antibody sequestration or PLC-mediated hydrolysis mimics the potentiating effects of bradykinin or NGF at the cellular level. Moreover, recruitment of PLC- γ to TrkA is essential for NGF-mediated potentiation of channel activity, and biochemical studies suggest that VR1 associates with this complex. These studies delineate a biochemical mechanism through which bradykinin and NGF produce hypersensitivity and might explain how the activation of PLC signalling systems regulates other members of the TRP channel family.

Nociceptors are specialized sensory neurons that detect chemical, physical or mechanical stimuli that could cause tissue damage. Inflammation increases the levels of extracellular protons, neurotransmitters and growth factors within the local tissue environment, thereby decreasing activation thresholds and enhancing the excitability of the nociceptor^{2,5}. Included among these agents are the nonapeptide bradykinin and the neurotrophin NGF^{2,6–8}. Sensitization by bradykinin or NGF might involve long-term changes in gene expression^{3,9,10}, but the ability of these factors to promote nociceptor sensitization within several minutes suggests that post-translational mechanisms are also involved. The downstream effector molecules that mediate sensitization by bradykinin or NGF have not been fully characterized³.

One candidate target on nociceptors is VR1, a cation channel that is activated by pungent vanilloid compounds (such as capsaicin), extracellular protons or noxious heat¹¹. This ability to integrate thermal and chemical stimuli makes VR1 potentially well suited to modulate sensitivity of the nociceptor after tissue injury¹². Indeed, VR1-deficient mice do not develop thermal hyperalgesia in response to inflammation^{13,14}. Because injury generates an array of chemical mediators, we asked whether bradykinin and NGF, specifically, contribute to VR1-dependent thermal hypersensitivity. We injected wild-type and VR1-deficient mice with bradykinin or NGF and measured paw withdrawal latencies from a radiant heat source before and after treatment. Each agent produced substantial sensitization in wild-type animals but not in VR1-null mice (Fig. 1),

demonstrating that VR1 is essential for the development of bradykinin- or NGF-induced thermal hypersensitivity *in vivo*.

Whereas extracellular protons interact directly with VR1 to potentiate its sensitivity to capsaicin or heat^{12,15}, NGF or bradykinin might modulate VR1 by binding to their own receptors on sensory neurons and activating second messenger signalling cascades. Indeed, VR1 belongs to the TRP channel family, some members of which are activated downstream of PLC-coupled receptors through as-yet undefined mechanisms¹⁶. To address this possibility, we expressed VR1 and BK₂ receptors together in human embryonic kidney (HEK293) cells and examined their sensitivity to VR1 agonists before and after exposure to bradykinin. Moderately acidic solutions (pH 6.4) evoked small but discernible inward currents at negative holding potentials. Subsequent treatment with bradykinin (20 nM) dramatically increased these responses ~25-fold (Fig. 2a). Even at physiological pH, bradykinin produced a slight increase in holding current that was attributable to VR1, based on characteristic outward rectification and inhibition by the vanilloid receptor antagonist capsazepine (see Supplementary Information). At a low dose of capsaicin (10 nM), bradykinin also produced a substantial (approximately fivefold), prolonged potentiation of evoked currents (Fig. 2b). In all cases, modulation was observed only in cells expressing both VR1 and BK₂ receptors.

To determine whether potentiation involves a change in channel gating, we analysed the kinetic properties of capsaicin-evoked

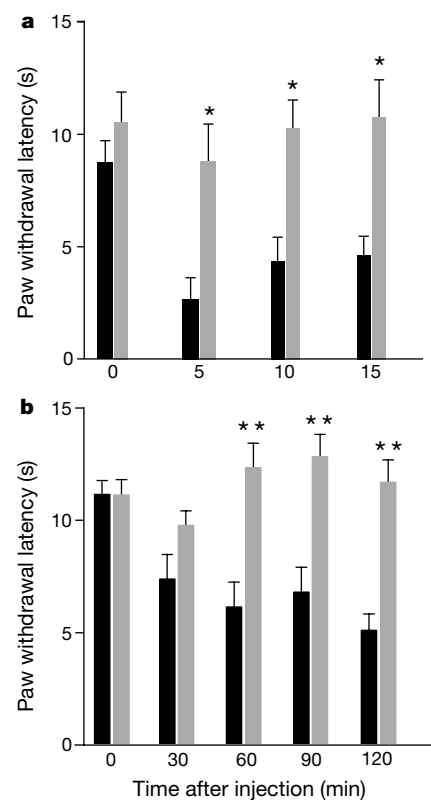


Figure 1 VR1 is essential for the development of bradykinin- or NGF-evoked thermal hypersensitivity *in vivo*. **a**, Wild-type (black) or VR1^{-/-} (grey) mice were injected intraplantarly with bradykinin and the response latency to radiant heating of the hind paw was measured at various times after injection (zero time point shows pre-injection response latency). *: $P < 0.05$ for VR1^{+/+} versus VR1^{-/-}; $P = 0.02$ for entire time course; $n = 7$ mice per genotype. **b**, NGF-evoked thermal hypersensitivity was measured as above. **: $P < 0.02$ for VR1^{+/+} versus VR1^{-/-}; $P = 0.01$ for the entire time course; $n = 7$ mice per genotype. NGF was administered subcutaneously and delivery to the receptive field (that is, the paw) might account for the slower onset of hyperalgesia relative to bradykinin. Neither group showed significant change in response latencies following saline injection (not shown).

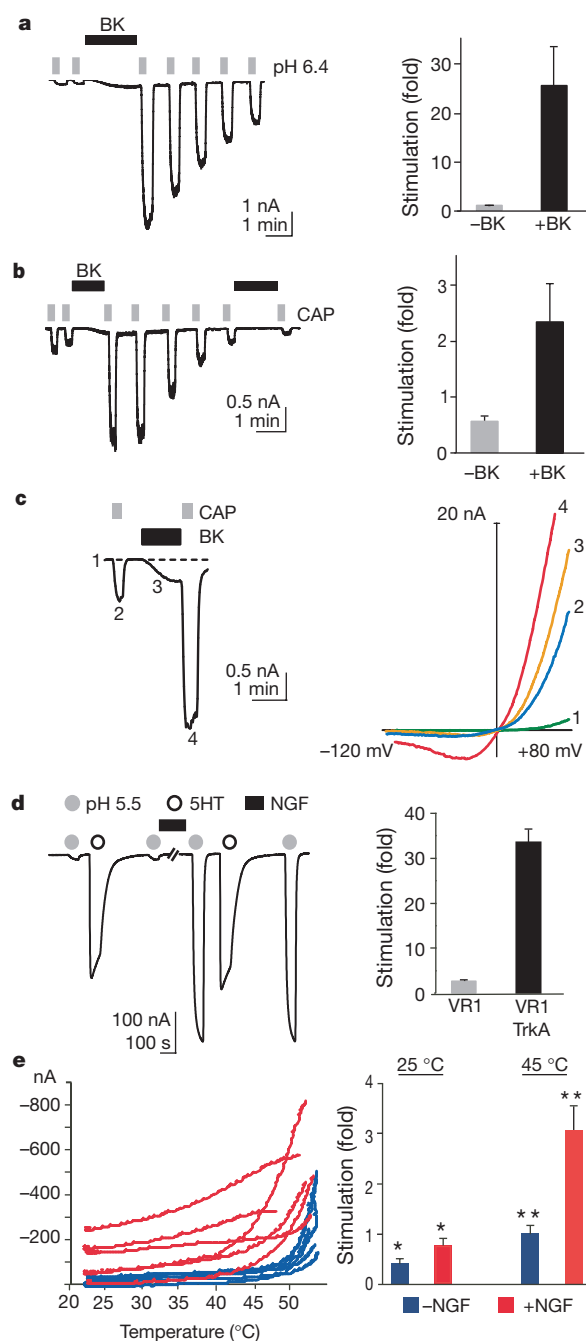


Figure 2 Bradykinin (BK) and nerve growth factor (NGF) sensitize VR1 in heterologous expression systems. **a**, Left: BK (20 nM) potentiation of proton-evoked (pH 6.4) responses in a voltage-clamped VR1-expressing HEK293 cell. Right: normalized proton response before (grey) and after (black) BK application ($n = 8$, $P < 0.05$). **b**, Left: BK potentiation of capsaicin-evoked (CAP; 10 nM) responses in a voltage-clamped VR1-expressing HEK293 cell. Right: capsaicin response normalized to a previous application before (grey) and after (black) BK treatment ($n = 16$, $P < 0.05$). **c**, Current–voltage relationships of VR1 responses to capsaicin (10 nM) before and after BK application reveal a change in gating. Numbers indicate points at which voltage-ramp traces (right) were obtained from the current trace (left). **d**, Left: NGF sensitizes proton-evoked currents in oocytes expressing VR1, TrkA, p75 and 5HT₃R-A. Break indicates 10 min NGF (0.75 nM) application. Protons (pH 5.5) and serotonin (5HT, 10 μ M) were perfused for 30 s. Right: averaged potentiation by NGF in these cells (black) versus oocytes expressing VR1 alone (grey; $n = 10$, $P < 0.001$). **e**, Heat-evoked currents in oocytes co-expressing TrkA. Cells were incubated with NGF (red, 0.75 nM) or vehicle; blue; ND-96 buffer, pH 7.5) for 10 min at room temperature before analysis ($n = 14$ – 16 ; *, $P < 0.05$; **, $P < 0.001$). Water-injected oocytes did not respond to heat (not shown). Averaged values (right) were normalized to current responses of vehicle-treated cells at 45 °C.

currents before and after bradykinin exposure. In the absence of any stimulus, VR1 displayed slight basal channel activity characterized by strong outward rectification; the inward current at negative holding potential was negligible (Fig. 2c, curve 1). Application of capsaicin produced a greater proportional increase in currents at negative membrane potentials, resulting in weaker outward rectification (Fig. 2c, curve 2). Upon activation of BK₂ receptors, capsaicine-inhibitable basal currents grew slowly over 2–3 min and rectification gradually weakened (Fig. 2c, curve 3). A second challenge with capsaicin further weakened the outward rectification, preferentially augmenting the inward component (Fig. 2c, curve 4). A similar phenomenon was observed when protons were used as agonist (not shown). These observations suggest that potentiation of VR1 by bradykinin primarily involves an alteration in channel gating.

Previous studies have shown that exposing cultured dorsal root ganglion (DRG) neurons to NGF for 10 min produces acute sensitization of capsaicin-evoked responses¹⁷. Similarly, we found that NGF produced robust (~30-fold) increases in proton-evoked

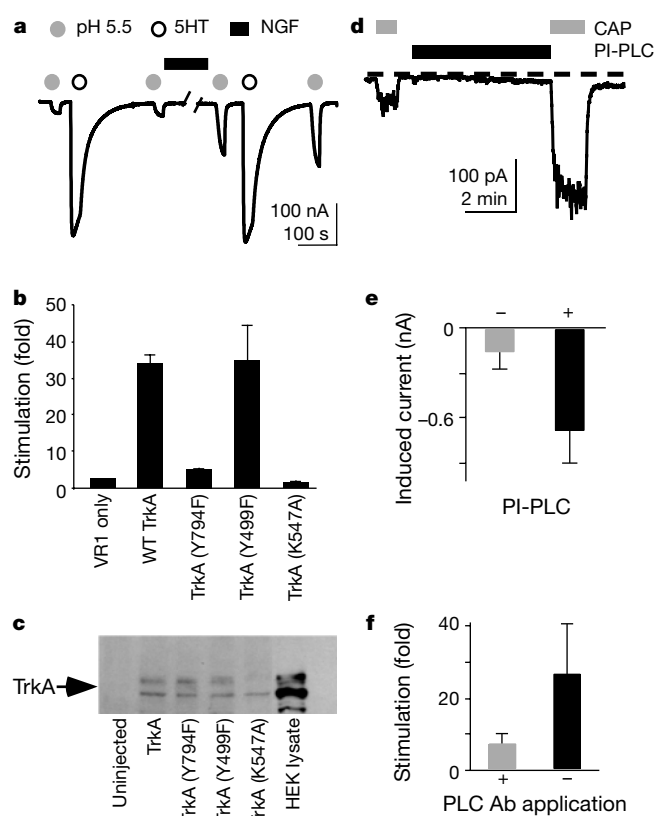


Figure 3 Phospholipase C is involved in nerve growth factor (NGF) modulation of VR1 function. **a**, Proton-evoked (pH 5.5) responses in a voltage-clamped oocyte expressing VR1, TrkA (Y794F), p75 and 5HT₃R-A. Analysis was performed as described in Fig. 2 ($P < 0.001$ versus cells expressing wild-type TrkA; $n = 10$). Break indicates 10 min NGF (0.75 nM) application. Protons (pH 5.5) and serotonin (5HT) (10 μ M) were perfused for 30 s. **b**, Histogram of the average potentiation of proton-evoked responses in oocytes co-expressing VR1 and wild-type or mutant TrkA receptors. **c**, Western blot of oocyte extracts indicates similar levels of wild type and mutant TrkA receptor expression. Lysate from TrkA-transfected HEK293 cells provides a control for TrkA immunoreactivity. **d**, **e**, Capsaicin-evoked (20 nM) currents before (grey) and after (black) application of inositol-phospholipid-specific phospholipase C (PI-PLC; 0.5 U ml⁻¹) to inside-out membrane patches excised from HEK293 cells expressing VR1 (–60 mV holding potential; $n = 12$, $P < 0.01$). **f**, PI-PLC potentiation (black) was reduced on pretreatment of PI-PLC with the anti-PLC antibody 72-24 (PLCAb; grey). Current was normalized to first capsaicin application ($n = 7$, $P < 0.05$).

(pH 5.5) currents in oocytes expressing both TrkA and VR1 (Fig. 2d). Potentiation was specific for VR1 because we found no change in the activity of co-expressed 5HT₃ serotonin-gated channels. Moreover, NGF had no effect unless TrkA was present (Fig. 2d). We also compared temperature response profiles for NGF-treated and untreated cells, and found that NGF produced a substantial decrease in the thermal threshold for VR1 activation, such that significant currents were observed even at room temperature (Fig. 2e). This was accompanied by an increase in the magnitude of responses at temperatures both below and above the threshold normally required for VR1 activation (~43 °C). These effects, as well as those described above for bradykinin, are reminiscent of the signature changes in nociceptor excitability produced by inflammation, including development of ongoing activity and sensitization to thermal stimuli¹⁰. Although the neurotrophin receptor p75 was usually expressed together with TrkA in these experiments, no difference was observed in its absence (not shown), which is consistent with recent evidence that p75 is not required for NGF-evoked thermal hypersensitivity *in vivo*¹⁸.

PLC stimulation is common to both NGF and bradykinin receptor activation, and we therefore asked whether this is an

obligate step in the VR1 potentiation process. A TrkA mutant (Y794F) that specifically abrogates recruitment of PLC- γ to the receptor complex¹⁹ showed greatly diminished NGF-mediated potentiation of proton-evoked currents ($14 \pm 7\%$ of the wild-type receptor), even though western blot analysis showed that mutant and wild-type receptors were expressed at comparable levels (Fig. 3a–c). By contrast, another TrkA mutant (Y499F) that disrupts coupling to the mitogen-activated-protein-kinase pathway¹⁹ was fully capable of potentiating VR1-evoked responses (Fig. 3a–c). These results demonstrate that PLC activation is the predominant pathway involved in VR1 potentiation by NGF. To determine whether PLC activity alone can potentiate VR1 responses, we applied a recombinant inositide-phospholipid-specific PLC (PI-PLC) directly to patches excised from VR1-expressing HEK293 cells. PI-PLC enhanced both basal channel activity and capsaicin-evoked responses, and these effects were significantly reduced by a neutralizing monoclonal antibody to the enzyme (Fig. 3d–f).

PLC catalyses the hydrolysis of membrane PtdIns(4,5)P₂ to yield inositol triphosphate and diacylglycerol, one consequence of which is activation of protein kinase C (PKC). Pharmacological studies suggest that bradykinin potentiates heat-evoked currents in sensory

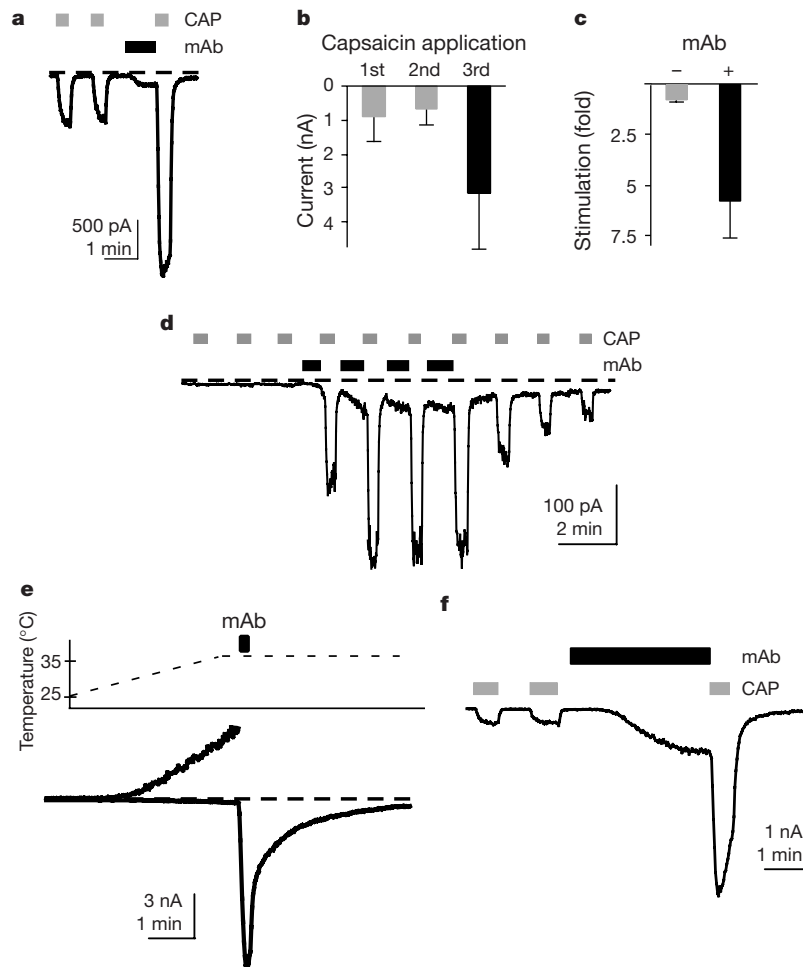


Figure 4 Phosphatidyl-4,5-inositol bisphosphate (PtdIns(4,5)P₂) antibody application mimics modulation of VR1 by bradykinin (BK) or nerve growth factor (NGF). **a**, Application of PtdIns(4,5)P₂ monoclonal antibody (mAb) to an inside-out membrane patch excised from VR1-expressing oocytes induced basal current and potentiated capsaicin (CAP, 20 nM)-evoked currents. **b**, Averaged response to capsaicin challenge in the absence (1st and 2nd; grey) and presence (3rd; black) of mAb. **c**, Current responses induced by capsaicin (normalized to previous application) showed more than sixfold potentiation by

mAb ($n = 12$, $P < 0.001$). **d**, Modulation by mAb was also observed in membrane patches excised from VR1-expressing HEK293 cells. **e**, Heat application to an excised membrane patch from a VR1-expressing HEK293 cell elicited a strongly rectifying basal current below 37 °C (upward deflection shows outward current at +70 mV holding potential before mAb application). Brief application of mAb then provoked a large inward current (–60 mV holding potential) at 37 °C. **f**, Potentiation of capsaicin-evoked (20 nM) responses by mAb was also observed in membrane patches excised from rat DRG neurons.

neurons through activation of PKC²⁰, and that PKC potentiates capsaicin-evoked responses in VR1-expressing oocytes²¹. However, we found that neither staurosporine nor Ro-31-8425, two PKC inhibitors, blocked NGF-elicited potentiation of VR1 in oocytes (not shown). In addition, attempts to mimic potentiation with the PKC activator phorbol 12-myristate 13-acetate (PMA, also known as TPA) were confounded by our observation that this agent antagonizes binding of [³H]-resiniferatoxin (a high-affinity capsaicin receptor agonist) to VR1, suggesting that PMA interacts directly with the channel (see Supplementary Information). Importantly,

we also found that PI-PLC potentiated VR1 currents in the absence of ATP.

Taken together, these results suggest that modulation of VR1 involves both PKC-dependent and PKC-independent mechanisms, with the latter possibly occurring as a direct consequence of PtdIns(4,5)P₂ hydrolysis. A number of ion channels are regulated in this manner, as exemplified by PtdIns(4,5)P₂ inhibition of cyclic nucleotide-gated channels²² or PtdIns(4,5)P₂ potentiation of G-protein-gated inwardly rectifying potassium channels^{23,24}. Anti-PtdIns(4,5)P₂ monoclonal antibodies sequester membrane PtdIns(4,5)P₂ and thus shift the sensitivity of these channels to cyclic GMP or G_{βγ}, respectively. When we applied this antibody to the cytoplasmic side of excised inside-out patches from VR1-expressing oocytes, we observed a gradual stimulation of basal channel activity (Fig. 4a–c) that displayed characteristic outward rectification. PtdIns(4,5)P₂ antibody also potentiated capsaicin-evoked currents, which normally display tachyphylaxis (a gradual reduction of induced current upon repetitive capsaicin application). Additionally, when the pipette solution was made more acidic (pH 6.4), PtdIns(4,5)P₂ antibody induced large inward currents (not shown). A similar induction was observed at neutral pH in cells expressing the VR1 mutants E600Q or E600K (not shown), which, under normal physiological conditions, behave like proton-potentiated wild-type channels¹⁵. PtdIns(4,5)P₂ antibody did not evoke responses in membrane patches from control oocytes (not shown). Taken together, these results demonstrate that PtdIns(4,5)P₂-antibody-activated currents are carried by VR1.

Cell types differ in their membrane lipid composition and we asked whether PtdIns(4,5)P₂ sequestration would also sensitize VR1 channels in mammalian cells, including sensory neurons. Very small inward currents were observed when 20 nM capsaicin was applied to inside-out membrane patches excised from VR1-transfected HEK293 cells. Application of PtdIns(4,5)P₂ antibody enhanced basal and capsaicin-induced currents. Both of these currents were suppressed by capsazepine (see Supplementary Information) and antibody effects were partially reversed on washout (Fig. 4d). This action was observed only when antibody was applied to the cytoplasmic side of the membrane (where PtdIns(4,5)P₂ resides) and a monoclonal antibody against progesterone had no effect (not shown). PtdIns(4,5)P₂ antibody also shifted the heat sensitivity of VR1 to a lower temperature range. When excised patches from these cells were exposed to a temperature ramp plateauing at 37 °C, a small VR1-mediated basal current developed. Brief application of antibody to this membrane patch then triggered a large inward current, indicating that the thermal sensitivity of VR1 can be modulated by PtdIns(4,5)P₂ in a membrane-delimited fashion (Fig. 4e). PtdIns(4,5)P₂ antibody also stimulated basal currents and sensitized capsaicin-evoked responses in rat DRG neurons (Fig. 4f). These results suggest that endogenous PtdIns(4,5)P₂ inhibits VR1 and that repression can be alleviated by agents that activate PLC.

Mammalian TRPC3 channels associate with TrkB receptors in the brain and these channels can be stimulated by brain-derived neurotrophic factor through a TrkB/PLC-dependent signalling pathway²⁵. We therefore asked whether VR1 and TrkA associate to form an analogous signalling complex. Immunoprecipitates prepared from HEK293 cells co-expressing VR1 and TrkA contained both proteins, irrespective of whether VR1 or TrkA antiserum was used to form the precipitate (Fig. 5a). Moreover, endogenous PLC-γ was detected in these samples, demonstrating that VR1, TrkA and PLC-γ can associate to form a ternary complex. Significantly less PLC-γ was recovered in immunoprecipitates from cells producing TrkA (Y794F) mutant receptors (Fig. 5a), which is consistent with our electrophysiology results showing reduced NGF potentiation of VR1 currents in oocytes producing this construct. In addition, VR1 associated with a catalytically inactive TrkA (K547A) mutant (see Supplementary Information), demonstrating that receptor kinase

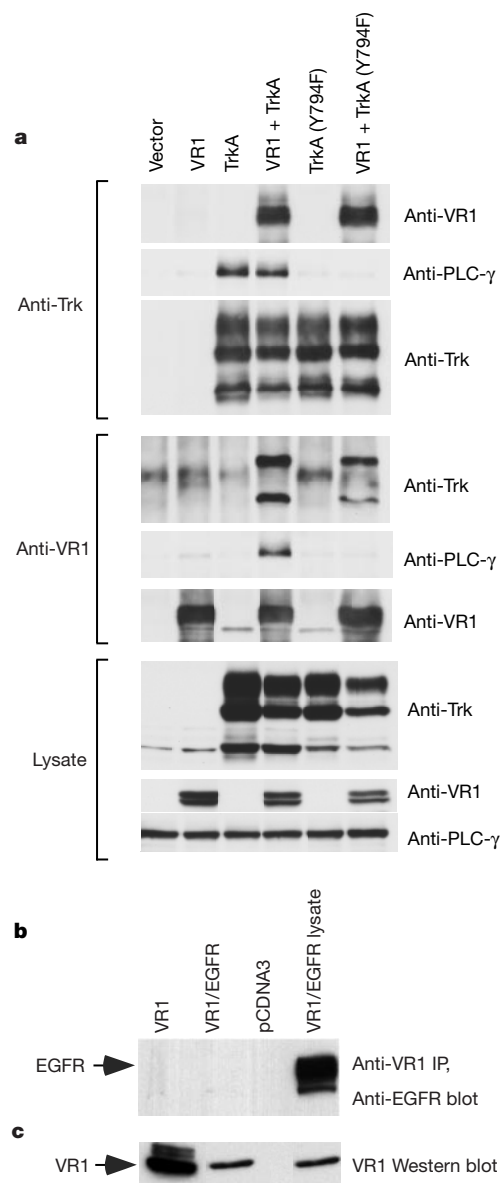


Figure 5 VR1, TrkA and phospholipase C-γ (PLC-γ) form a signalling complex. **a**, HEK293 cells were transfected with the indicated cDNAs (top). Immunoprecipitates were formed with the antibodies indicated on the right and visualized on western blots with the antibodies indicated on the left. Total lysates were also analysed to demonstrate equivalent protein expression among samples. TrkA is typically produced as three differently glycosylated isoforms migrating as species with relative molecular masses of 140,000, 110,000 and 80,000 (*M_r* 140K, 110K and 80K, respectively). **b**, VR1 does not form a complex with the epidermal growth factor receptor (EGFR). HEK293 cells were transfected with indicated cDNAs and immune complexes formed using VR1 antiserum. Blots were probed with EGFR antibody. The rightmost lane contains lysate from cells producing both VR1 and EGFR. **c**, VR1 expression in total cell lysates.

activity is not a prerequisite for interaction. Although we have not shown that TrkA and VR1 associate in sensory neurons, the interaction that we observe in HEK293 cells is specific because the epidermal growth factor (EGF) receptor, a closely related receptor tyrosine kinase, does not immunoprecipitate with VR1 when expressed in these cells (Fig. 5b, c).

A subset of vertebrate and invertebrate TRP channels are activated by neurotransmitter, hormone or growth-factor receptors that couple to PLC signalling pathways. How PLC activation leads to TRP channel opening is still the subject of significant debate, but a number of mechanisms have been proposed that involve calcium store depletion, direct interaction with inositol triphosphate receptors or activation by lipid metabolites produced downstream of PLC¹⁶. Our data suggest that VR1 can be activated or potentiated at a very early point in the pathway as an immediate consequence of PLC activation and PtdIns(4,5)P₂ hydrolysis. Indeed, application of G_{α11} to inside-out membrane patches activates *Drosophila* TRPL channels through a PLC-mediated mechanism¹⁶, and application of PtdIns(4,5)P₂ partially inhibits TRPL opening²⁶. The ability of exogenously applied anandamide or diacylglycerol to activate VR1 or other TRP channels in a PKC-independent manner^{27,28} raises the possibility that these lipid messengers mediate their effects by displacing PtdIns(4,5)P₂ or other membrane lipids from an inhibitory site on the channel complex.

In the *Drosophila* eye, TRP channels exist in a complex with other components of the phototransduction machinery, an arrangement that influences the sensitivity and kinetics of the signalling process^{29,30}. Whether a similar signalling scaffold exists in nociceptors remains to be determined. Interestingly, we have found that a variety of PLC-coupled receptors modulate VR1 function in heterologous systems (not shown). Moreover, preliminary experiments suggest that EGF receptor activation also sensitizes VR1, albeit to a lesser extent (approximately fivefold) than that typically observed with NGF (see Supplementary Information). Thus, whereas the formation of a specific protein complex is not a prerequisite for modulation, complex formation might facilitate the process. Activation of different PLC-coupled receptors by individual components of the inflammatory milieu might allow for spatially and quantitatively distinct mechanisms of nociceptor sensitization. □

Methods

Behaviour

Bradykinin (20 µl of a 10 nM solution in 0.9% saline) was injected intraplantarly into one hind paw of wild-type or VR1^{-/-} mice (adult males, ~25 g body weight) and response latencies to a radiant heat source were measured as described¹³. NGF (2.5 S, gift of W. Mobley) was administered subcutaneously (1 mg kg⁻¹ in 0.9% saline). Significance was determined using repeated analysis of variance with Fisher's PLSD (protected least significant difference) *post-hoc* analysis.

Oocyte electrophysiology

Defolliculated stage VI *Xenopus laevis* oocytes were injected with 0.1–1.0 ng rat VR1, rat TrkA, rat p75, human EGFR or mouse 5HT₃R-A *in vitro* transcribed RNA. Oocytes were maintained in modified Barth's saline solution for 4–6 days before analysis. Two-electrode voltage-clamp analysis was performed using a Geneclamp 500 amplifier (Axon Instruments) and MacLab A/D converter (–45 mV holding potential). ND-96 (with Ba²⁺) recording solution was buffered with 10 mM Na-MES (pH 5.5) or 10 mM Na-HEPES (pH 7.5) and bath chamber perfused at ~1 ml min⁻¹. Temperature ramps were generated as described¹². For patch-clamp experiments, 2–10 ng of VR1 cRNA were injected into oocytes and responses measured 2–5 weeks after injection. Recombinant rat NGF-β, human EGF (Sigma) and bradykinin (Bachem) were diluted into the perfusate.

Mammalian cell electrophysiology

HEK293 cells were cultured in DMEM/F12 with 10% FBS (fetal bovine serum) and transfected using Lipofectamine, Superfect or Eugene 6 according to manufacturers' protocols. Cells were transfected with 400 ng plasmid DNA encoding rat VR1, with or without 400 ng human BK₂ receptor complementary DNA. To identify transfected cells, an enhanced green fluorescence protein reporter plasmid was also transfected at one-tenth the concentration of receptor cDNAs. Cells were plated onto coverslips 1–3 days before recording and examined 4–7 days after transfection. Dissociated neurons from rat or mouse DRG were prepared as described¹³. Voltage-clamp experiments were performed at –60 mV holding potential with 320 ms voltage ramp from –120 mV to +80 mV at 1 Hz.

Data were acquired using pClamp (Axon Instruments) or Pulse-Pulsefit (HEKA GmbH) software. Recordings were filtered at 5 kHz and sampled at 1 kHz. Standard bath solution for patch-clamp experiments contained 10 mM Tris/HCl, 1 mM EGTA, 1 mM MgCl₂ and 150 mM CsCl at pH 7.4. The pH 6.4 solution contained 20 mM citric acid and 1 mM MgCl₂, titrated with CsOH to pH 6.4 and supplemented with CsCl to make the final Cs⁺ concentration 150 mM. For excised patch experiments, standard bath solution was used in both the pipette and perfusion solution. No ATP was added to the solutions. Diameters of patch pipettes were 20–50 µm for oocyte recordings and 12–15 µm for HEK293 cells or neurons. For whole-cell recording, a 0.9× standard bath solution supplemented with 0.24 mM CaCl₂, 0.2 mM Na₂GTP and 0.3 mM Na₂ATP was used in the pipette, adjusted with Tris-base to a final pH of 7.3. Pipette resistance in this solution was typically 2–3 MΩ. PI-PLC and its neutralizing antibody were from Molecular Probes, and PtdIns(4,5)P₂ and progesterone antibodies from Perseptive Biosystem. Titres of PtdIns(4,5)P₂ antibody used in individual experiments were 1:200 (Fig. 4a–c, e), 1:2,000 (Fig. 4d) and 1:500 (Fig. 4f). A perfusion stage with heating resistors was used to generate a heat ramp plateauing at 37 °C, and the bath temperature was monitored with probes (Warner Instruments).

Biochemistry

HEK293 cells were maintained in DMEM containing 10% FBS. Cells were plated at 70–80% confluency and subjected to calcium phosphate or lipofectamine transfection with approximately equal amounts of plasmid DNAs. At 24–48 h after transfection, cells were collected and lysed in 1 ml TNE buffer (10 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% NP40) containing 0.12 mg ml⁻¹ phenylmethyl sulphonyl fluoride, 2 µg ml⁻¹ leupeptin, 1 µg ml⁻¹ aprotinin, 10 mM NaF and 1 mM Na₃VO₄ on ice. Lysates of equivalent protein content were incubated overnight at 4 °C with anti-pan-Trk polyclonal antibody C14 (Santa Cruz Biotech) (1.5 µg) or anti-VR1 antiserum¹² (1:1,000). Immune complexes were immobilized on protein-A-Sepharose beads, washed with ice-cold TNE, boiled in SDS sample buffer, separated by SDS-PAGE and transferred onto Immobilon-P (Millipore) membranes. Membranes were blocked in TBST buffer (20 mM Tris pH 7.5, 500 mM NaCl, 0.1% Tween-20) containing 5% non-fat milk, then incubated for 2 h at room temperature or overnight at 4 °C in TBST with 1% milk and one of the following primary antibodies: anti-Trk 45 antiserum (1:2,000); anti-VR1 antiserum (1:2,000); or anti-PLC-γ antibody (Upstate Biotech) (0.7 µg ml⁻¹). Membranes were washed with TBST, then incubated with horseradish-peroxidase-conjugated goat anti-rabbit or goat anti-mouse secondary antibodies. Immunoreactive protein bands were detected by enhanced chemiluminescence using ECL reagents. For western blot analysis of oocyte proteins, 10 cells were lysed in 100 µl of 100 mM NaCl, 1% Triton X-100, 20 mM Tris-HCl pH 7.6 and complete protease inhibitor cocktail (Roche) by drawing 5–10 times each through 15 gauge and 26 gauge needles. Lysates were centrifuged 10 min at 15,000g and supernatant fractions were passed twice through 0.22 µm Spin-X filters (Costar). Final flow through (~20 µl) was denatured in SDS sample buffer and analysed as above.

Received 8 February; accepted 18 April 2001.

- McMahon, S. B. & Bennett, D. L. H. in *Textbook of Pain* (eds Wall, P. D. & Melzack, R.) 105–128 (Harcourt, London, 1999).
- Bevan, S. in *Textbook of Pain* (eds Wall, P. D. & Melzack, R.) 85–103 (Harcourt, London, 1999).
- Ganju, P., O'Bryan, J. P., Der, C., Winter, J. & James, I. F. Differential regulation of SHC proteins by nerve growth factor in sensory neurons and PC12 cells. *Eur. J. Neurosci.* **10**, 1995–2008 (1998).
- Burgess, G. M., Mullaney, I., McNeill, M., Dunn, P. M. & Rang, H. P. Second messengers involved in the mechanism of action of bradykinin in sensory neurons in culture. *J. Neurosci.* **9**, 3314–3325 (1989).
- Woolf, C. J. & Salter, M. W. Neuronal plasticity: increasing the gain in pain. *Science* **288**, 1765–1769 (2000).
- Shu, X. Q. & Mendell, L. M. Neurotrophins and hyperalgesia. *Proc. Natl Acad. Sci. USA* **96**, 7693–7696 (1999).
- Nicholas, R. S., Winter, J., Wren, P., Bergmann, R. & Woolf, C. J. Peripheral inflammation increases the capsaicin sensitivity of dorsal root ganglion neurons in a nerve growth factor-dependent manner. *Neuroscience* **91**, 1425–1433 (1999).
- Koltzenburg, M., Bennett, D. L., Shelton, D. L. & McMahon, S. B. Neutralization of endogenous NGF prevents the sensitization of nociceptors supplying inflamed skin. *Eur. J. Neurosci.* **11**, 1698–1704 (1999).
- Woolf, C. J. Phenotypic modification of primary sensory neurons: the role of nerve growth factor in the production of persistent pain. *Phil. Trans. R. Soc. Lond. B* **351**, 441–448 (1996).
- Koltzenburg, M. The changing sensitivity in the life of the nociceptor. *Pain* **6** (Suppl.), S93–S102 (1999).
- Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
- Tominaga, M. *et al.* The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron* **21**, 531–543 (1998).
- Caterina, M. J. *et al.* Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**, 306–313 (2000).
- Davis, J. B. *et al.* Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. *Nature* **405**, 183–187 (2000).
- Jordt, S. E., Tominaga, M. & Julius, D. Acid potentiation of the capsaicin receptor determined by a key extracellular site. *Proc. Natl Acad. Sci. USA* **97**, 8134–8139 (2000).
- Harteneck, C., Plant, T. D. & Schultz, G. From worm to man: three subfamilies of TRP channels. *Trends Neurosci.* **23**, 159–166 (2000).
- Shu, X. & Mendell, L. M. Nerve growth factor acutely sensitizes the response of adult rat sensory neurons to capsaicin. *Neurosci. Lett.* **274**, 159–162 (1999).
- Bergmann, L., Reiter, R., Toyka, K. V. & Koltzenburg, M. Nerve growth factor evokes hyperalgesia in mice lacking the low-affinity neurotrophin receptor p75. *Neurosci. Lett.* **255**, 87–90 (1998).
- Stephens, R. M. *et al.* Trk receptors use redundant signal transduction pathways involving SHC and PLC-γ1 to mediate NGF responses. *Neuron* **12**, 691–705 (1994).

20. Cesare, P., Dekker, L. V., Sardini, A., Parker, P. J. & McNaughton, P. A. Specific involvement of PKC- ϵ in sensitization of the neuronal response to painful heat. *Neuron* **23**, 617–624 (1999).
21. Premkumar, L. S. & Ahern, G. P. Induction of vanilloid receptor channel activity by protein kinase C. *Nature* **408**, 985–990 (2000).
22. Womack, K. B. *et al.* Do phosphatidylinositides modulate vertebrate phototransduction? *J. Neurosci.* **20**, 2792–2799 (2000).
23. Huang, C. L., Feng, S. & Hilgemann, D. W. Direct activation of inward rectifier potassium channels by PIP_2 and its stabilization by $\text{G}_{\beta\gamma}$. *Nature* **391**, 803–806 (1998).
24. Zhang, H., He, C., Yan, X., Mirshahi, T. & Logothetis, D. E. Activation of inwardly rectifying K^+ channels by distinct $\text{PtdIns}(4,5)\text{P}_2$ interactions. *Nature Cell Biol.* **1**, 183–188 (1999).
25. Li, H. S., Xu, X. Z. & Montell, C. Activation of a TRPC3-dependent cation current through the neurotrophin BDNF. *Neuron* **24**, 261–273 (1999).
26. Estacion, M., Sinkins, W. G. & Schilling, W. P. Regulation of *Drosophila* transient receptor potential-like (TrpL) channels by phospholipase C-dependent mechanisms. *J. Physiol.* **530**, 1–19 (2001).
27. Zygmunt, P. M. *et al.* Vanilloid receptors on sensory nerves mediate the vasodilator action of anandamide. *Nature* **400**, 452–457 (1999).
28. Hofmann, T. *et al.* Direct activation of human TRPC6 and TRPC3 channels by diacylglycerol. *Nature* **397**, 259–263 (1999).
29. Chevesich, J., Kreuz, A. J. & Montell, C. Requirement for the PDZ domain protein, INAD, for localization of the TRP store-operated channel to a signaling complex. *Neuron* **18**, 95–105 (1997).
30. Scott, K. & Zuker, C. S. Assembly of the *Drosophila* phototransduction cascade into a signalling complex shapes elementary responses. *Nature* **395**, 805–808 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We are grateful to W. Neuhauser for providing dissociated DRG neurons, W. Mobley for 2.5 S NGF, D. Shelton for advice regarding NGF administration and J. Poblete and K. Simpson for technical assistance. We thank members of our laboratories for many helpful suggestions and constructive criticism. This work was supported by a NSF predoctoral fellowship (E.P.), NIH predoctoral neuroscience training grant (S.S.), a German Academy of Natural Scientists Leopoldina postdoctoral fellowship (S.-E.J.) and by the NIH (M.V.C., A.I.B. and D.J.).

Correspondence and requests for materials should be addressed to D.J. (e-mail: julius@socrates.ucsf.edu).

Agonist-independent activation of metabotropic glutamate receptors by the intracellular protein Homer

F. Ango*, L. Prézeau*, T. Muller†, J. C. Tu‡, B. Xiao‡, P. F. Worley‡, J. P. Pin*, J. Bockaert* & L. Fagni*

* CNRS-UPR 9023, CCIPE, 141 Rue de la Cardonille, 34000 Montpellier, France
 † BAYER AG, Pharma Research, 42096 Wuppertal, Germany
 ‡ Department of Neurosciences, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

G-protein-coupled receptors (GPCRs) transduce signals from extracellular transmitters to the inside of the cell by activating G proteins. Mutation and overexpression of these receptors have revealed that they can reach their active state even in the absence of agonist, as a result of a natural shift in the equilibrium between their inactive and active conformations¹. Such agonist-independent (constitutive) activity has been observed for the glutamate GPCRs (the metabotropic glutamate receptors mGluR1a and mGluR5) when they are overexpressed in heterologous cells². Here we show that in neurons, the constitutive activity of these receptors is controlled by Homer proteins, which bind directly to the receptors' carboxy-terminal intracellular domains^{3,4}. Disruption of this interaction by mutagenesis or antisense strategies, or expression of endogenous Homer1a (H1a), induces constitutive activity in mGluR1a or mGluR5. Our results show that these glutamate GPCRs can be directly activated by intracellular proteins as well as by agonists.

When expressed in HEK-293 cells, mGluR1a and mGluR5 display

constitutive activity that cannot be inhibited by competitive antagonists² or by the non-competitive and neutral mGluR1 antagonist CPCCOEt⁵. This basal activity can be inhibited only by the selective non-competitive antagonists (so-called inverse agonists) BAY 36-7620 (ref. 6) and MPEP⁷, respectively. In cultured cerebellar granule cells, which naturally express mGluR1a^{2,8}, BAY 36-7620 reduces neither basal inositol phosphate formation (Table 1) nor the basal open probability (P_o) of the Ca^{2+} -dependent big K^+ (BK) channel (Fig. 1a, b). These two pathways are activated by mGluR1/5 agonists^{9–12}, including DHPG⁸, the effect of which is antagonized by CPCCOEt (Table 1, Fig. 1a). Although cerebellar granule cells do not express native mGluR5 (refs 2, 8), overexpression of the cloned receptor (mGluR5a) can be obtained after transfection of these cells^{8,13}. Even under these conditions, treatment with MPEP decreases neither basal inositol phosphate formation (Table 1) nor basal BK channel activity (Fig. 2a), although the transfected mGluR5a was functionally expressed. It was fully activated by DHPG and this effect was blocked by the competitive neutral mGluR antagonist MCPG (Table 1, Fig. 2a, b). Thus, constitutive activity of native mGluR1a or overexpressed mGluR5a cannot be detected in cultured cerebellar granule cells.

When expressed in HEK-293 cells, the mGluR1a, 5a and 5b splice variants display much higher agonist-independent activity than mGluR1b, 1c and 1d (ref. 2). Interestingly, mGluR1a, 5a and 5b, but not mGluR1b, 1c and 1d, interact with the intracellular Homer proteins^{4,14}. We investigated the role of Homer in the activity of mGluR1a and mGluR5a in cultured cerebellar granule cells. These

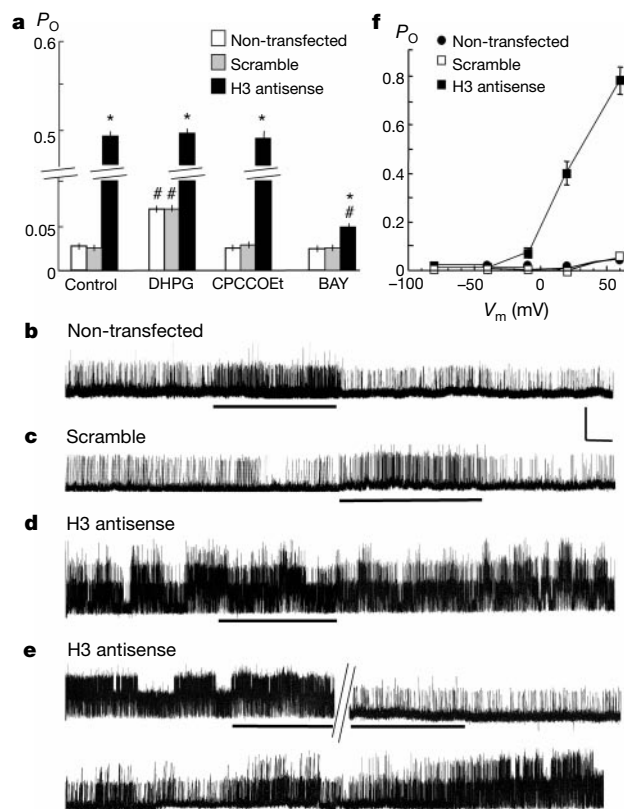


Figure 1 Knock-down of H3 results in constitutive mGluR1a activity. Data obtained from non-transfected neurons or neurons transfected with the indicated oligonucleotides. **a**, BK channel P_o measured in the absence (control) or presence of the indicated drugs. Asterisks and hash symbols, significantly different from non-transfected or control neurons, respectively ($P < 0.01$). **b–e**, BK channel recordings. Horizontal bars represent DHPG (**b–d**) or BAY 36-7620 (**e**; time lag, 5 min) applications. Calibration bars (5 pA, 10 s) apply to **b–e**. **f**, BK channel P_o -voltage relationships (mean $\pm$ s.e.m.; $n \approx 10$).

Table 1 Knock-down of H3 or loss of mGluR5–Homer interaction increases inositol phosphate formation

	Absence of drug	CPCCOEt (250 μ M)	BAY 36-7620 (10 μ M)	DHPG (400 μ M)	MPEP (10 μ M)	MCPG (500 μ M)	DHPG + MPEP	DHPG + MCPG
Non-transfected	100 ¹	100 $\pm$ 2	101 $\pm$ 3	200 $\pm$ 5*	103 $\pm$ 4	102 $\pm$ 3	101 $\pm$ 3	101 $\pm$ 3
Scramble	100 $\pm$ 2	100 $\pm$ 3	102 $\pm$ 4	200 $\pm$ 7*	NT	NT	NT	NT
H3 antisense	150 $\pm$ 6*	150 $\pm$ 4*	110 $\pm$ 3	200 $\pm$ 5*	NT	NT	NT	NT
mGluR5a wild type	100 ²	101 $\pm$ 4	102 $\pm$ 3	239 $\pm$ 9*	102 $\pm$ 4	100 $\pm$ 3	101 $\pm$ 5	102 $\pm$ 3
mGluR5a P1127A	100 $\pm$ 4	NT	NT	250 $\pm$ 9*	NT	NT	NT	NT
mGluR5a F1128R	170 $\pm$ 8*	NT	NT	260 $\pm$ 12*	102 $\pm$ 5	170 $\pm$ 7*	105 $\pm$ 4	175 $\pm$ 8*

Each value (mean $\pm$ s.e.m. of three experiments performed in triplicate) is the percentage of inositol phosphate formation obtained in the absence of drug (second column), or presence of the indicated compounds (following columns). Each line is obtained in non-transfected cultures, or cultures transfected with the indicated oligonucleotides or mGluR5a complementary DNAs (in the presence of CPCCOEt). Refs 1 and 2 are the reference values for the first and last three lines, respectively.

*Significantly different from reference ($P < 0.001$).

cells express Homer3 (H3; Fig. 3a, b, d) and mGluR1a (Fig. 3e)^{8,15}. Transfection of fluorescent H3 antisense RNA resulted in knock-down of this protein (Fig. 3c, d) and revealed constitutive inositol phosphate formation (Table 1) and high spontaneous BK channel activity (Fig. 1a, d, f). These effects did not result from release of endogenous glutamate, as they were unaffected by CPCCOEt (Fig. 1a, Table 1) but were completely blocked by BAY 36-7620 (Fig. 1a, e, Table 1). This effect of BAY 36-7620 resulted from blockade of constitutively active endogenous mGluR1a, rather than BK channels, as these channels were recorded in the cell-attached configuration. Moreover, the drug did not affect the basal activity of BK channels in non-transfected neurons, or neurons transfected with scramble oligonucleotide (Fig. 1a). We verified that H3 knock-down did not affect expression of endogenous mGluR1a (Fig. 3e). Also, no knock-down of H3 was observed in neurons transfected with scrambled oligonucleotide (Fig. 3b, d), and the endogenous mGluR1a was not constitutively active although it was still responsive to DHPG (Table 1, Fig. 1a and c). Together, these results indicate that H3 prevents agonist-independent activity of mGluR1a.

We investigated whether constitutive activity of mGluR resulted from a lack of specific interaction between the receptor and H3. The mGluR5a F1128R mutant does not interact with Homer proteins¹⁴.

Cultured cerebellar granule cells transfected with this mutant displayed elevated spontaneous BK channel activity (Fig. 2a, d) and inositol phosphate formation (Table 1). These activities were inhibited by the inverse agonist MPEP (Fig. 2a, e), but not by the neutral antagonist, MCPG (Fig. 2a), indicating that the mGluR5a F1128R mutant was constitutively active. This effect probably did not result from increased expression of the mutant receptor, as there was no difference in expression of the mutant and wild-type receptors in the soma¹³. Furthermore, cerebellar granule cells transfected with the mGluR5a P1127A mutant, which interacts with Homer proteins¹⁴, displayed normal and MPEP-insensitive basal BK channel activity (Fig. 2a, c). These results showed that disruption of the mGluR5–H3 interaction in neurons resulted in constitutive receptor activity. Unexpectedly, such disruption of the mGluR1a/5–H3 interaction enhanced BK channel activity more effectively than DHPG in control cells (Figs 1a and 2a). It is possible that ‘chronic’ activity of the receptor induces biochemical changes that lead to higher amplification of the transduction pathway than acute stimulation by the agonist. Whatever its origin, the increased basal BK channel activity was completely blocked by the inverse agonist, showing that this effect was entirely receptor-dependent.

Unlike constitutively expressed H3, H1a is the product of an

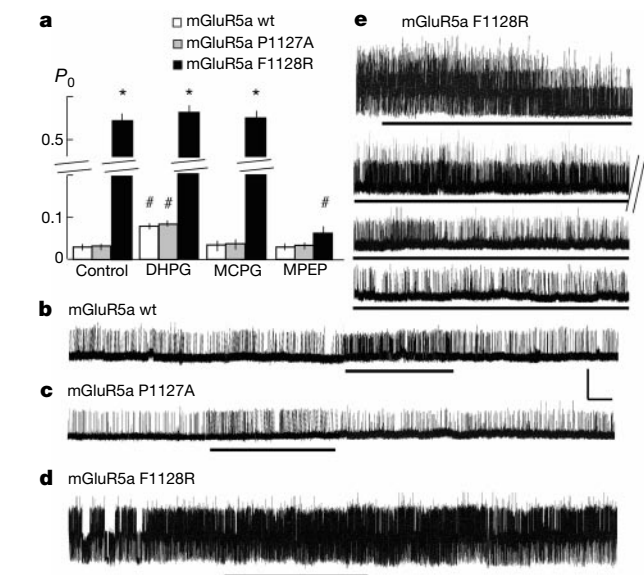


Figure 2 Interaction between mGluR5a and H3 prevents constitutive receptor activity. **a**, BK channel P_0 obtained in neurons transfected with wild type (wt) or the indicated mGluR5a mutants, in the absence (control) or presence of the indicated drugs. Asterisks and hash symbols, significantly different from wild-type mGluR5a or control, respectively ($P < 0.01$). **b–e**, BK channel recordings from neurons transfected with the indicated mGluR5a. Horizontal bars represent DHPG (**b–d**) or MPEP (**e**) applications. In **e**, time lag is 5 min. Calibration bars (5 pA, 10 s) apply to **b–e**. All experiments performed in the presence of CPCCOEt.

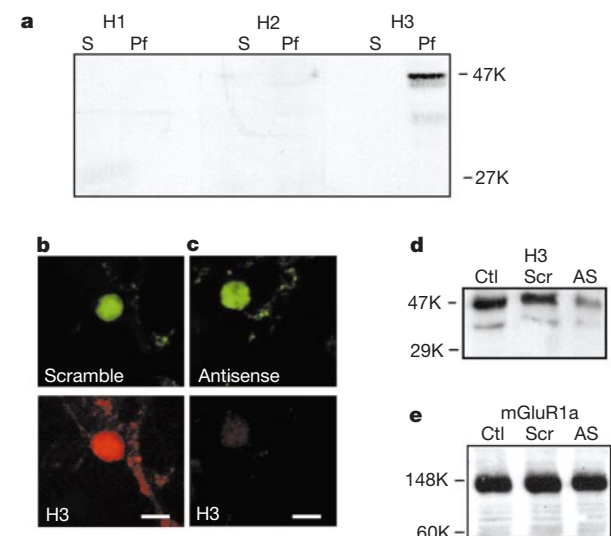


Figure 3 Knock-down of H3. **a**, Western blot from a non-transfected culture, using the indicated anti-Homer antibodies. Calibration marks correspond to apparent relative molecular mass of H1a (27,000) and H1b, 1c, 2 and 3 (47,000). S and Pf, soluble and particulate fractions, respectively. **b**, Fluorescent signals from transfected scramble oligonucleotide (upper panel) and H3 protein immunostaining (lower panel), in the same neuron. **c**, As in **b** but with transfected H3 antisense oligonucleotide. Calibration bars, 10 μ m. **d**, **e**, Immunoblots from non-transfected cultures (Ctl) or cultures transfected with scramble (Scr) or H3 antisense (AS) oligonucleotides, revealed using an anti-H3 (**d**) or anti-mGluR1a (**e**) antibody.

immediate-early gene^{16,17}. This protein is a dominant-negative regulator of the interaction between the other Homer proteins and mGluR1a or mGluR5 (refs 4, 14). We induced expression of endogenous H1a in cultured cerebellar granule cells to disrupt interaction between the endogenous mGluR1a and H3 by exposing the neurons for 30 min to a mixture of the ionotropic glutamate receptor agonists NMDA (*N*-methyl-D-aspartate) and kainate¹³. Twelve hours after the treatment, the neurons expressed H1a (Fig. 4a), and this protein could interact directly with the endogenous mGluR1a. Indeed, mGluR1 was co-immunoprecipitated with H1a and, in parallel, a smaller amount of mGluR1a was co-immunoprecipitated with H3 (Fig. 4d). These results revealed direct competition between endogenous H1a and H3 on mGluR1a. These neurons also displayed high spontaneous BK channel activity that was inhibited by BAY 36-7620 (Fig. 4b), but not by CPCCOEt (not shown). This result indicated that induced H1a triggered constitutive activity of mGluR1a. Similar results were obtained in H1a-transfected neurons (Fig. 4b). Immunoblots performed on biotinylated and non-biotinylated membrane proteins showed that this effect did not result from increased cell-surface expression of mGluR1a (Fig. 4c).

We have shown that constitutive activity of mGluR1a and mGluR5a, in cultured neurons, was kept undetectable by interaction between the receptors and native H3. These findings, with the fact that the shorter variants, mGluR1b, 1c and 1d, are virtually devoid of constitutive activity², indicate that the extended C termini of mGluR1a, 5a and 5b may be important for constitutive activity. Also, constitutive activity of GPCRs has most often been observed in transfected cell lines, where receptor-interacting proteins might be absent. We found, however, that the constitutive mGluR1a activity was not significantly altered by co-transfection of H3 in HEK-293 cells (data not shown), indicating that, in addition to Homer, other

partners present in neurons but not in HEK-293 cells may be required to prevent constitutive activity of the receptor. The Shank and MAGUK proteins are likely candidates, as they form complex assemblies with Homer at the postsynaptic density¹⁸.

Neuronal excitation and expression of endogenous H1a disrupt interaction between mGluR1a and H3, revealing constitutive activity of the receptor. Our data show that neuronal activity-dependent modulation of the interaction between a GPCR and its associated intracellular protein can elicit receptor activity independently of agonist binding, and with a time constant different from that of agonist-mediated activation. Such a situation may exist under conditions of H1a induction, that is, during synaptic plasticity or convulsive seizures^{16,17}. Our results are consistent with previous work showing that H1a has no effect on membrane localization of mGluR1a and mGluR5 (ref. 19). This argues against the possibility that constitutive activity of mGluR1a or mGluR5 resulted from alterations in receptor membrane targeting by H1a. Alternatively, as for other GPCRs^{20–24}, allosteric conformational changes in the C termini of mGluR1a and mGluR5 could result in constitutive activity, as this domain interacts with Homer proteins and participates in G-protein activation²⁵. Our results indicate that such a conformational change may depend on complex receptor–protein interactions that create constraints in the C terminus of GPCRs and alter the functional coupling to the G protein. □

Methods

Primary culture transfection and immunocytochemistry

Cerebellar cultures were prepared from one-week-old mice and transfected as described⁸. The antisense oligonucleotide (GGCCTGTTCCTGGCTGTGGA; Eurogentec) corresponded to positions 1–21 of the mouse H3 gene. These antisense and scramble oligonucleotides were phosphorothioated and labelled with S-modified [3']fluorescein. In electrophysiological experiments, the mGluR5a and H1a expression plasmids were co-transfected with the reporter gene, green fluorescent protein.

We performed immunocytochemistry and western blots on 7–10-day *in vitro* cultures using rabbit anti-H1, -H2 and -H3 antibodies⁴ (1:1,000), as described¹³, and the polyclonal anti-mGluR1a antibody (1:1,000; Upstate). For mGluR1a co-immunoprecipitation, lysates prepared from untreated culture or culture treated with NMDA and kainate were precipitated with anti-H3 or anti-H1 antibody and revealed with a mouse anti-mGluR1a antibody (1:2,000; PharMingen), as described⁴. Biotinylation of cell-surface proteins was performed on living cultures as described²⁶.

Patch-clamp recordings and inositol phosphate measurements

We measured single BK channel P_o from cell-attached patches, at a membrane potential of +20 mV, under conditions previously described⁸. The P_o values represented cumulative channel open probabilities divided by the number of open levels. Determination of inositol phosphate formation was as described²⁷. All results were expressed as mean $\pm$ s.e.m. of at least 10 experiments and analysed using the Student's *t*-test.

Received 17 November 2000; accepted 25 April 2001.

- Kenakin, T. Agonist-specific conformations. *Trends Pharmacol. Sci.* **18**, 416–417 (1997).
- Prezeau, L. *et al.* Changes in the carboxyl-terminal domain of metabotropic glutamate receptor 1 by alternative splicing generate receptors with differing agonist-independent activity. *Mol. Pharmacol.* **49**, 422–429 (1996).
- Kato, A. *et al.* Novel members of Ves/Homer family of PDZ proteins that bind metabotropic glutamate receptors. *J. Biol. Chem.* **273**, 23969–23975 (1998).
- Xiao, B. *et al.* Homer regulates the association of group I metabotropic glutamate receptors with multivalent complexes of Homer-related, synaptic proteins. *Neuron* **21**, 707–716 (1998).
- Litschig, S. *et al.* CPCCOEt, a noncompetitive metabotropic glutamate receptor 1 antagonist, inhibits receptor signaling without affecting glutamate binding. *Mol. Pharmacol.* **55**, 453–461 (1999).
- Carroll, F. Y. *et al.* BAY36-7620: A potent non-competitive mGlu1 receptor antagonist with inverse agonist activity. *Mol. Pharmacol.* **59**, 965–973 (2001).
- Pagano, A. *et al.* The non-competitive antagonists MPEP and CPCCOEt interact with overlapping binding pockets in the transmembrane region of group-I metabotropic glutamate receptors. *J. Biol. Chem.* **275**, 33750–33758 (2000).
- Ango, F. *et al.* A simple method to transfer plasmid DNA into neuronal primary cultures: functional expression of the mGlu5 receptor in cerebellar granule cells. *Neuropharmacology* **38**, 793–803 (1999).
- Sladczek, F., Pin, J. P., Récasens, M., Bockaert, J. & Weiss, S. Glutamate stimulates inositol phosphate formation in striatal neurons. *Nature* **317**, 717–719 (1985).
- Nicoletti, F. *et al.* Coupling of inositol phospholipid metabolism with excitatory amino acid recognition sites in rat hippocampus. *J. Neurochem.* **46**, 40–46 (1986).
- Chavis, P., Ango, F., Michel, J. M., Bockaert, J. & Fagni, L. Modulation of big K⁺ channel activity by ryanodine receptors and L-type Ca²⁺ channels in neurons. *Eur. J. Neurosci.* **10**, 2322–2327 (1998).
- Fagni, L., Bossu, J. L. & Bockaert, J. Activation of a large-conductance Ca²⁺-dependent K⁺ channel by stimulation of glutamate phosphoinositide-coupled receptors in cultured cerebellar granule cells. *Eur. J. Neurosci.* **3**, 778–789 (1991).

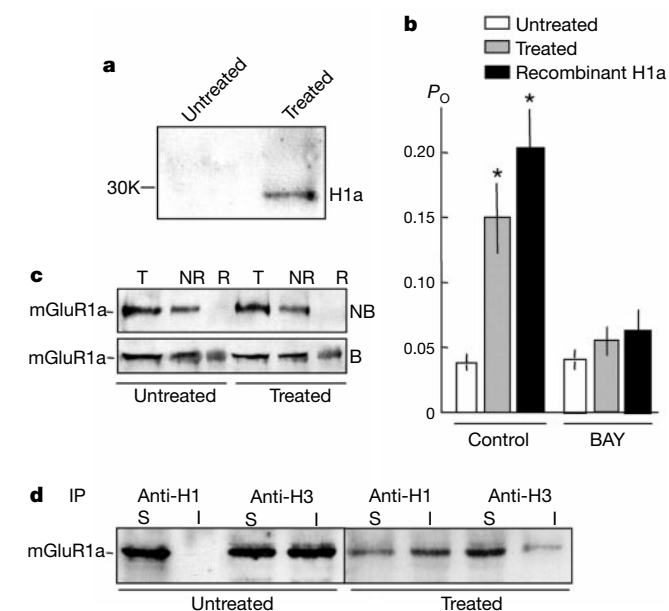


Figure 4 Endogenous or recombinant H1a induces constitutive mGluR1a activity. Data obtained from untreated cultures or cultures treated with NMDA and kainate (both at 100 μ M). **a**, Western blot of particulate fraction using an anti-H1 antibody. The revealed band corresponds to H1a. **b**, BK channel basal P_o in the absence (Ctl) or presence of BAY 36-7620. Asterisks, significantly different from untreated ($P < 0.05$). **c**, Biotinylated (B) or non-biotinylated surface proteins (NB) were separated on streptavidine beads. The retained (R), not retained (NR) and total (T) fractions were analysed in western blots with an anti-mGluR1a antibody. **d**, Neuronal membranes were solubilized and immunoprecipitated (IP) with anti-H1 or anti-H3 antibodies. Soluble (S) and immunoprecipitated (I) fractions were analysed in western blots for the presence of mGluR1a.

13. Ango, F. *et al.* Dendritic and axonal targeting of type 5 metabotropic glutamate receptor is regulated by Homer1 proteins and neuronal excitation. *J. Neurosci.* **20**, 8710–8716 (2000).
14. Tu, J. C. *et al.* Homer binds a novel proline-rich motif and links group I metabotropic glutamate receptors with IP3 receptors. *Neuron* **21**, 717–726 (1998).
15. Prézeau, L. *et al.* Pharmacological characterization of metabotropic glutamate receptors in several types of brain cells in primary cultures. *Mol. Pharmacol.* **45**, 570–577 (1994).
16. Brakeman, P. R. *et al.* Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**, 284–288 (1997).
17. Kato, K., Ozawa, F., Saitoh, Y., Hairi, K. & Inokuchi, K. *vesl*, a gene encoding VASP/Ena family related protein, is upregulated during seizure, long-term potentiation and synaptogenesis. *FEBS Lett.* **412**, 183–189 (1997).
18. Tu, J. C. *et al.* Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins. *Neuron* **23**, 583–592 (1999).
19. Roche, K. W. *et al.* Homer 1b regulates the trafficking of group I metabotropic glutamate receptors. *J. Biol. Chem.* **274**, 25953–25957 (1999).
20. Claeysen, S. *et al.* Constitutively active mutants of 5-HT4 receptors: are they in unique active states? *EMBO Rep.* **2**, 61–67 (2001).
21. Parker, E. M. & Ross, E. M. Truncation of the extended carboxyl-terminal domain increases the expression and regulatory activity of the avian β -adrenergic receptor. *J. Biol. Chem.* **266**, 9987–9996 (1991).
22. Matus-Leibovitch, N., Nussensveig, D. R., Gershengorn, M. C. & Oron, Y. Truncation of the thyrotropin-releasing hormone receptor carboxyl tail causes constitutive activity and leads to impaired responsiveness in *Xenopus* oocytes and AtT20 cells. *J. Biol. Chem.* **270**, 1041–1047 (1995).
23. Hasegawa, H., Negishi, M. & Ichikawa, A. Two isoforms of the prostaglandin E receptor EP3 subtype different in agonist-independent constitutive activity. *J. Biol. Chem.* **271**, 1857–1860 (1996).
24. Claeysen, S., Sebben, M., Becamel, C., Bockaert, J. & Dumuis, A. Novel brain-specific 5-HT4 receptor splice variants show marked constitutive activity: role of the C-terminal intracellular domain. *Mol. Pharmacol.* **55**, 910–920 (1999).
25. Mary, S., Gomez, J., Prézeau, L., Bockaert, J. & Pin, J.-P. A cluster of basic residues in the carboxy-terminal tail of the short mGluR1 variants impairs their coupling to PLC. *J. Biol. Chem.* **273**, 425–432 (1998).
26. Galvez, T. *et al.* Mapping the agonist-binding site of GABAB type 1 subunit sheds light on the activation process of GABAB receptors. *J. Biol. Chem.* **275**, 41166–41174 (2000).
27. Joly, C. *et al.* Molecular, functional and pharmacological characterization of the metabotropic glutamate receptor type 5 splice variants: comparison with mGluR1. *J. Neurosci.* **15**, 3970–3981 (1995).

Acknowledgements

We thank F. Gasparini and R. Kuhn (Novartis Parma) for providing MPEP; D. Robbe for preliminary experiments on mGluR5 mutants; and J. Perroy, C. de Cole and C. Becamel for help in improving our transfection method and experiments with HEK-293 cells and western blots. This work was supported by grants from CNRS, Bayer (France) and FRM. L.F. was also supported by AFM, J.B. by GIP-HMR, and J.P.P. by Retina France and the French Ministry. F.A. holds a grant from Sanofi-Synthelabo (France).

Correspondence and requests for materials should be addressed to L.F. (e-mail: fagni@bacchus.montp.inserm.fr).

Nodal signalling in the epiblast patterns the early mouse embryo

Jane Brennan*, Cindy C. Lu*, Dominic P. Norris*, Tristan A. Rodriguez†, Rosa S. P. Beddington† & Elizabeth J. Robertson*

* Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

† Division of Mammalian Development, National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK

Shortly after implantation the mouse embryo comprises three tissue layers. The founder tissue of the embryo proper, the epiblast, forms a radially symmetric cup of epithelial cells that grows in close apposition to the extra-embryonic ectoderm and the visceral endoderm. This simple cylindrical structure exhibits a distinct molecular pattern along its proximal–distal axis¹. The anterior–posterior axis of the embryo is positioned later by coordinated cell movements that rotate the pre-existing proximal–distal axis^{2–5}. The transforming growth factor- β family member Nodal is known to be required for formation of the anterior–posterior axis⁶. Here we show that signals from the

epiblast are responsible for the initiation of proximal–distal polarity. Nodal acts to promote posterior cell fates in the epiblast and to maintain molecular pattern in the adjacent extra-embryonic ectoderm. Both of these functions are independent of Smad2. Moreover, Nodal signals from the epiblast also pattern the visceral endoderm by activating the Smad2-dependent pathway required for specification of anterior identity in overlying epiblast cells. Our experiments show that proximal–distal and subsequent anterior–posterior polarity of the pregastrulation embryo result from reciprocal cell–cell interactions between the epiblast and the two extra-embryonic tissues.

The site of primitive-streak formation during gastrulation marks the future posterior side of the embryo and thus defines the anterior–posterior axis. Recent experiments indicate that the epiblast acquires molecular asymmetry before overt streak induction¹. Thus, epiblast cells on the prospective anterior side are instructed to maintain an anterior fate by signals from the anterior visceral endoderm (AVE). *Nodal* is expressed throughout the epiblast and the overlying visceral endoderm at early post-implantation stages⁷. As development proceeds, a proximal–distal gradient of *Nodal* expression formed within the epiblast rapidly resolves to the future posterior side where the primitive streak will form. Embryos lacking Nodal activity arrest shortly before gastrulation and fail to establish a primitive streak^{8,9}. *Smad2*, an effector of Nodal signalling¹⁰, is also essential for establishment of the anterior–posterior axis^{6,11}. It was important to compare the onset of patterning abnormalities in embryos lacking *Nodal* or *Smad2*. As shown in Fig. 1, *Wnt3* (ref. 12) and *Eomesodermin* (*Eomes*)¹³ are initially expressed radially in the proximal epiblast and rapidly become confined to the posterior side of the embryo, consistent with their essential roles during mesoderm formation. Surprisingly, *Nodal*-deficient embryos lack posterior markers including *Wnt3*, *Eomes* (Fig. 1b, n), *T* and *Fgf8* (data not shown). Similarly, *Nodal* mutants failed to express *Cripto*³ (Fig. 1e), a marker normally expressed throughout the epiblast of wild-type embryos before regressing to the posterior side (Fig. 1d), and *Lefty2* (ref. 14), a member of the transforming growth factor- β (TGF β) family (Fig. 1q) initially expressed at gastrulation (Fig. 1p). The striking absence of any posterior markers is not simply caused by a general block to differentiation because *Nodal* mutants rapidly downregulate expression of the pluripotential marker gene *Pou1f1* (*Oct3/4*)¹⁵ (Fig. 1h). Surprisingly, we observed the reciprocal expression pattern of *Wnt3*, *Cripto* and *Eomes* in *Smad2*-deficient embryos (Fig. 1c, f, o), such that all three genes are uniformly expressed at high levels throughout the epiblast. By contrast, expression of *Lefty2* is greatly reduced in *Smad2*-deficient embryos (Fig. 1r). *Lefty2* acts as an antagonist to inhibit mesoderm formation¹⁴. Thus loss of *Lefty2* may contribute to the widespread expression of posterior markers and exclusive formation of extra-embryonic mesoderm in *Smad2*-deficient embryos.

We investigated whether these defects in epiblast patterning are associated with abnormal development of the extra-embryonic ectoderm (ExE) by examining the expression of two distal markers, *Bmp4* (ref. 11) and *Eomes* (Fig. 1i–o). Expression of both genes is transiently detected in *Nodal* mutants (Fig. 1j and n, and data not shown). As judged by hybridization with a probe for *L14* (ref. 16) as well as histological criteria, the ExE develops normally in 6.5 days post coitum (d.p.c.) *Nodal* mutants (Fig. 2k', see also Supplementary Information). Thus, this progressive loss of signal does not simply reflect selective tissue degeneration. Rather, continued Nodal signalling in the epiblast is required for maintaining molecular pattern in the adjacent ExE. Previous studies demonstrate that *Smad2*-deficient embryos express *Nodal* abundantly throughout the epiblast¹¹. Interestingly, *Bmp4* and *Eomes* (Fig. 1o) expression domains are maintained in *Smad2*-mutant embryos. Thus, Nodal signals that pattern the ExE are mediated through *Smad2*-independent pathways.

Considerable data indicate that signals from the distal-most visceral endoderm cells, fated to form the AVE, normally antagonize signals received by the proximal epiblast and thus act to maintain anterior identity¹. The discrete population of visceral endoderm overlying the distal end of the epiblast selectively expresses *Hex*¹⁷, *Lefty1* (refs 14, 18) and *Cer1* (ref. 19), whereas *Nodal*², *Otx2* (ref. 20), *Lhx1* (*Lim1*) and *Foxa2* (*Hnf3β*) (ref. 18) are broadly expressed throughout this tissue layer. Interestingly, in *Smad2* (ref. 11) or in *Lhx1/Foxa2* double mutants¹⁸ the AVE fails to form and the entire epiblast adopts a proximal character. We used a *lacZ* transgene regulated by *Hex* enhancer sequences active in the visceral endoderm (T.A.R. and R.S.P.B., unpublished results), as well as a panel of gene markers, to assess the extent of visceral endoderm patterning. Overall, we observed that *Nodal* mutants phenocopy *Smad2*-deficient embryos by failing to establish molecular pattern within the visceral endoderm (Fig. 2a–k and data not shown). As a control for the integrity of the visceral endoderm these mutant embryos were shown to express abundant levels of H19 (ref. 21) (data not shown). *Otx2* is robustly expressed in both the visceral endoderm and epiblast (Fig. 2j). Previous experiments have shown that *Otx2* activity in the visceral endoderm at early stages is necessary to establish normal anterior–posterior pattern, whereas *Otx2* expression in the epiblast at later stages is required for continued anterior development⁴. Interestingly, *Nodal* mutants fail to express *Otx2* transcripts in the visceral endoderm, and *Otx2* is only weakly and transiently expressed at very early stages in the epiblast (Fig. 2k). Loss of *Otx2* activity in the visceral endoderm probably disrupts reciprocal tissue interactions required to maintain *Otx2* expression

and anterior fates within the epiblast. *Foxa2* expression in wild-type embryos is restricted to a discrete population of anterior streak cells located several cell diameters away from the ExE (Fig. 2l). Unexpectedly, *Foxa2*, which is not activated in either the epiblast or visceral endoderm of *Smad2* mutants (data not shown), is found ectopically expressed throughout the epiblast of *Nodal* mutants (Fig. 2m). Deregulated *Foxa2* expression might simply reflect a general epiblast-patterning defect. However, we found that *Gsc*²², a marker normally co-expressed with *Foxa2* in the anterior streak, was absent in *Nodal* mutants (data not shown). Thus the loss of *Nodal* signalling probably disrupts interactions between the epiblast and the ExE that normally inhibit *Foxa2* expression in the proximal epiblast.

We and others^{23,24} have previously identified two discrete regulatory elements in the *Nodal* locus, one lying 5' to the coding region and the other within the first intron, that collectively direct *Nodal* expression during early development (Fig. 3; D.P.N. and E.J.R., unpublished observations). Sequences lying within the *Nodal* intronic enhancer are recognized by the forkhead DNA-binding protein FAST (*foxh1/2*), which associates with activated *Smad2/3/4* complexes formed in response to activin/*Nodal* signals^{25,26}. Thus the intronic enhancer could potentially regulate aspects of *Nodal* expression by an autoregulatory mechanism. To test this hypothesis, we assessed *Nodal* transcription in a *Nodal*-deficient background using the *Nodal*^{lacZ} reporter allele²⁷. As expected in heterozygous embryos, *Nodal*^{lacZ} expression detected throughout the epiblast and the visceral endoderm gradually becomes restricted to the posterior side of the epiblast (Fig. 3a). In contrast, in *Nodal*^{lacZ} homozygous

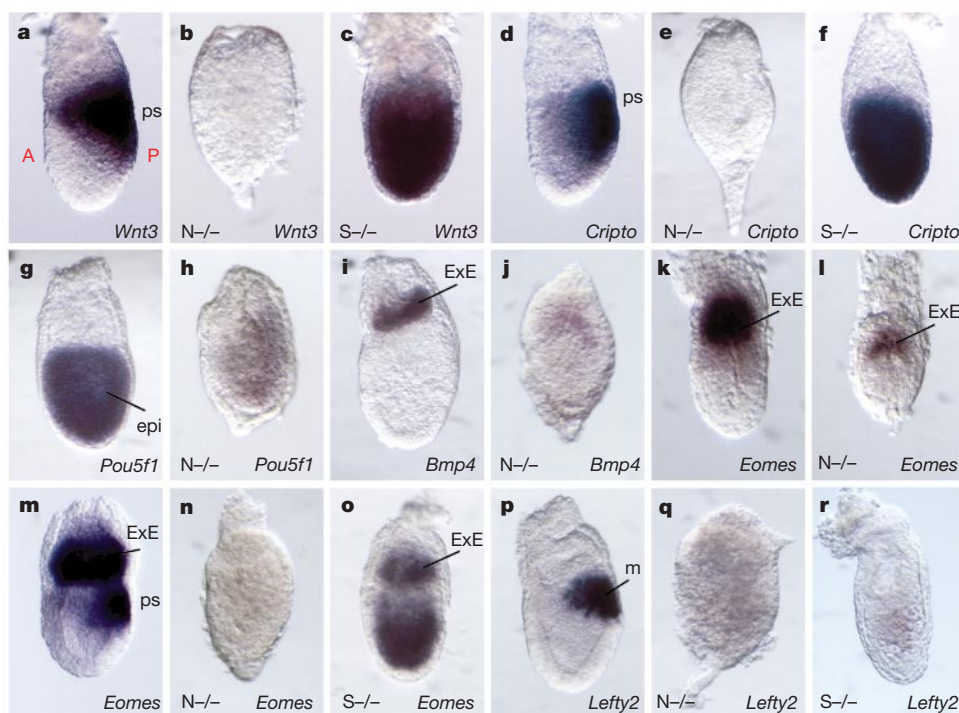


Figure 1 Distinct patterning defects exhibited by epiblast tissue deficient in *Nodal* or *Smad2*. Whole-mount *in situ* hybridization analysis of wild-type (a, d, g, i, k, m, p), *Nodal*-mutant (*N*^{−/−}) (b, e, h, j, l, n, q) and *Smad2*-mutant (*S*^{−/−}) (c, f, o, r) embryos at early gastrulation. Lateral views of embryos are shown at 6.5 d (a–j, m–r) and 5.5 d (k, l), with anterior (A) to the left and posterior (P) to the right. Expression of *Wnt3* and *Cripto* is restricted to the proximal posterior region of the wild-type embryo (a, d). *Nodal* mutants lack expression of *Wnt3* and *Cripto* (b, e), yet *Smad2* mutants express *Wnt3* and *Cripto* throughout their epiblast (c, f). The pluripotential marker *Pou5f1* is abundantly expressed in wild-type (g) but drastically downregulated in *Nodal*-mutant (h) embryos. Compared with the ExE of wild-type embryos (i), the *Bmp4* expression domain is smaller and less robust in

Nodal-mutant (j) embryos. At 5.5 d, *Eomesodermin* (*Eomes*) is expressed in the ExE of wild-type embryos (k) and *Nodal* mutants (l). One day later, the *Nodal* mutant (n) loses expression, whereas the wild-type embryo maintains *Eomes* expression in the ExE and also in the posterior epiblast (m). *Lefty2* is expressed in the anterior half of the primitive streak and in the mesoderm, exiting through it (p). In contrast, *Lefty2* is not expressed in *Nodal* mutants (q) and only weakly expressed throughout the epiblast of *Smad2* mutants (r). Embryos were collected from intercrosses established between heterozygous *Nodal*^{lacZ}/+ mice and between heterozygous *Smad2*^{Robm1}/+ mice. *In situ* hybridization was performed as previously described¹¹. epi, epiblast; ExE, extra-embryonic ectoderm; m, mesoderm; ps, primitive streak.

mutants, the reporter is transiently activated in small patches within the proximal epiblast (Fig. 3b). Moreover the *Nodal*^{lacZ} signal is lost by 7.0 d.p.c. (data not shown), and homozygous *Nodal*^{lacZ} mutants entirely lack expression in the visceral endoderm (Fig. 3b'). These data suggest that Nodal activity in the epiblast is required for activation of the locus in the visceral endoderm by means of the intronic enhancer. To test this possibility, we examined expression of a *lacZ* transgene carrying the intronic enhancer in various

contexts. The transgene is activated in both the epiblast and the visceral endoderm of wild-type embryos²⁴ (Fig. 3d). However, we failed to detect expression in either the epiblast or visceral endoderm of *Nodal*-mutant embryos (Fig. 3e). Thus, activation of this enhancer is strictly dependent on Nodal signalling. These observations suggest that Nodal expression in the proximal epiblast is activated by the 5' enhancer (Fig. 3b), whereas the intronic enhancer acts in an autoregulatory manner to amplify *Nodal*

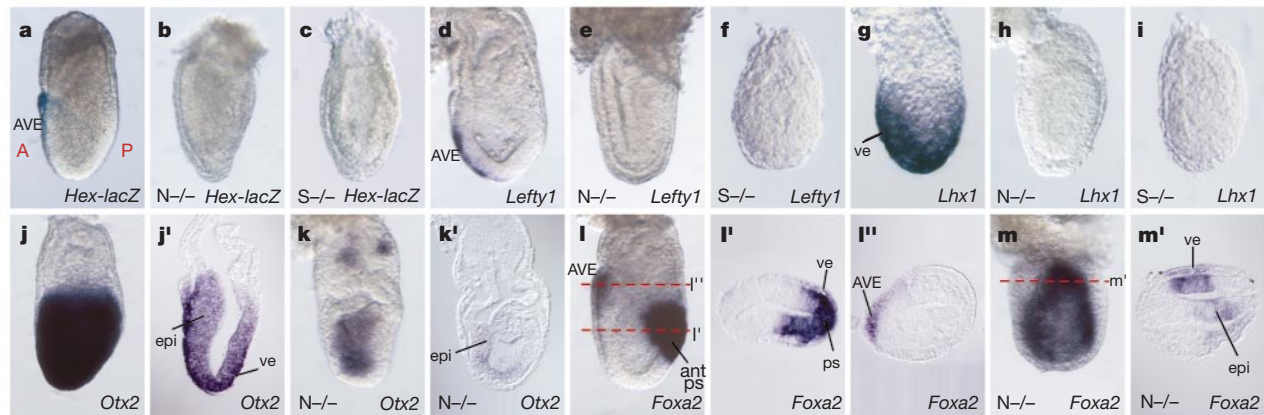


Figure 2 Visceral endoderm deficient in *Nodal* or *Smad2* lacks anterior identity. β gal (a–c) and whole-mount *in situ* hybridization (d–m) analysis of wild-type (a, d, g, j, l), *Nodal*-mutant (N–/–) (b, e, h, k, m) and *Smad2*-mutant (S–/–) (c, f, i) embryos at early gastrulation. Lateral views are shown, with anterior (A) to the left and posterior (P) to the right. Sections are longitudinal (j', k') and transverse (l', l'', m'). The *Hex*–*lacZ* transgene and *Lefty1*, activated in AVE of wild-type embryos (a, d), are not expressed in *Nodal* (b, e) or *Smad2* (c, f) mutants. *Lhx1*, normally present throughout the visceral endoderm before gastrulation (g), is also absent from *Nodal* and *Smad2* mutants (h, i). *Otx2* is normally expressed throughout the epiblast (j) and the overlying visceral endoderm

(j'). In *Nodal* mutants *Otx2* is expressed in the epiblast (k) but not in the visceral endoderm (k'). *Foxa2* is normally expressed in the anterior primitive streak, visceral endoderm overlying the streak (l') and AVE (l''). *Nodal* mutants express *Foxa2* ectopically throughout the epiblast (m) and the overlying visceral endoderm (m'). Whole-mount *in situ* hybridization was conducted as for Fig. 1. The *Hex*–*lacZ* transgene was constructed by cloning a 3.5-kb *Sac*II–*Spe*I intronic fragment from the *Hex* locus into a cassette containing *lacZ* under the control of a minimal promoter for the human β -globin gene (T.A.R. and R.S.P.B., unpublished results). ant, anterior; epi, epiblast; m, mesoderm; ps, primitive streak; ve, visceral endoderm.

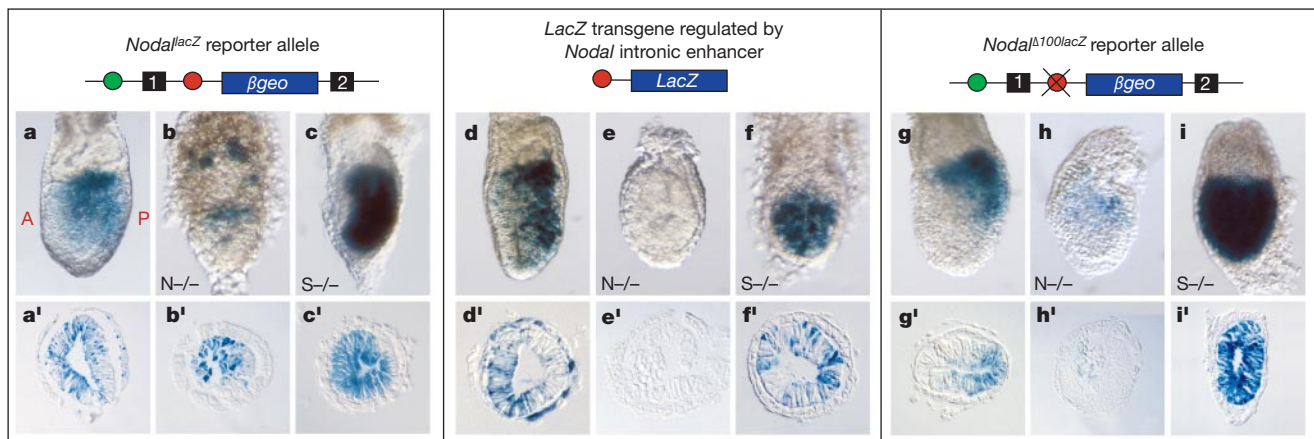


Figure 3 Tissue-specific *Smad2*-dependent and -independent signals regulate Nodal activities in the early embryo. β gal activity in early gastrulation control (a, d, g), *Nodal*-mutant (b, e, h) and *Smad2*-mutant (c, f, i) embryos. Lateral view (a–i) and transverse sections (a'–i') with anterior (A) to the left and posterior (P) to the right. *Nodal* regulatory elements acting at early stages are depicted as a green circle (5' early epiblast enhancer) and a red circle (intronic enhancer). In control embryos, *Nodal*^{lacZ} is expressed in the proximal posterior epiblast and overlying visceral endoderm (a). In contrast, *Nodal*-deficient embryos express the reporter allele in the proximal epiblast but not the visceral endoderm (b). In the absence of *Smad2*, the *Nodal*^{lacZ} reporter, which is expressed robustly throughout the epiblast, is similarly absent from the visceral endoderm (c) (the non-expressing tissue to the left of the *Smad2*-mutant epiblast is mis-localized ExE¹¹). The intronic enhancer normally activates *lacZ* in both the proximal posterior epiblast and overlying visceral endoderm (d). However, embryos lacking *Nodal* fail to express the

transgene (e). *Smad2*-deficient embryos express *lacZ* in the epiblast but not the visceral endoderm (f). In control embryos, deletion of the intronic enhancer (red circle) fails to influence expression in the proximal epiblast but eliminates expression from the visceral endoderm (g). Loss of *Nodal* has no effect on the proximal expression domain although the levels are reduced (h). In the absence of *Smad2*, expression is detectable throughout the epiblast (i) in keeping with expanded expression domains of other posterior markers. Embryos were collected from the following crosses and stained for β gal activity as previously described¹¹: *Nodal*^{lacZ}/+ intercross (a, b); double-heterozygous male (*Nodal*^{lacZ}/+, *Smad2*^{Robm1}/+) crossed to *Smad2*^{Robm1}/+ female (c); 413d/+ male⁹ carrying intronic enhancer transgene crossed to 413d/+ female (d, e); *Smad2*^{Robm1}/+ male carrying intronic enhancer transgene crossed to *Smad2*^{Robm1}/+ female (f); *Nodal*^{lacZ}/+ male crossed to 413d female (g, h); and double-heterozygous male (*Smad2*^{Robm1}/+, *Nodal*^{lacZ}/+) crossed to *Smad2*^{Robm1}/+ female.

expression throughout the epiblast and overlying visceral endoderm. To test this possibility, we examined expression of the *Nodal*^{Δ100lacZ} allele, generated by germline deletion of the intronic enhancer in the context of the *Nodal*^{lacZ} locus²⁴. Removal of the intronic enhancer did not interfere with initial activation, but was

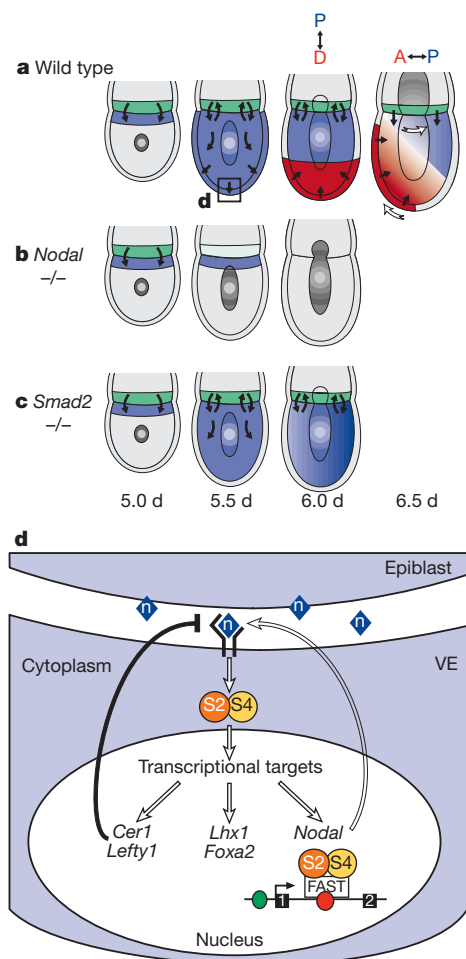


Figure 4 Temporally and spatially distinct Nodal activities pattern the anterior–posterior (A↔P) axis. **a**, *Nodal* is initially expressed in a proximal ring of epiblast in response to signals from the ExE. *Nodal* signalling maintains the expression of ExE markers and amplifies its own expression throughout the epiblast and into the overlying visceral endoderm (VE) through an autoregulatory mechanism. *Nodal* signalling in the epiblast activates the *Smad2* pathway in the visceral endoderm, leading to the expression of target genes that pattern the visceral endoderm (described in **d**). Signals from the visceral endoderm maintain the expression of genes such as *Otx2* in the distal epiblast while the proximal epiblast expresses mesoderm markers such as *Cripto*, *Wnt3* and *Eomes*. Proximal–distal (P↔D) polarity in gene expression patterns is converted to anterior–posterior polarity by the movement of distal endoderm to the prospective anterior side and of proximal epiblast to the posterior side. **b**, In *Nodal* mutants, ExE markers are initially expressed but fail to be maintained. In the absence of *Nodal*, mesoderm markers are not expressed and the embryo arrests development as an unpatterned egg cylinder. **c**, The initial expression and activity of *Nodal* in the proximal epiblast is unaffected in *Smad2* mutants, allowing the expression of a panel of mesoderm-inducing genes. As the *Smad2* pathway is critical for global endoderm patterning as well as for the formation of the AVE, loss of *Smad2* eliminates reciprocal signalling that normally restricts expression of markers such as *Wnt3* to the prospective posterior side. This results in the deregulated expression of posterior markers throughout the entire epiblast. Green shows expression of ExE markers such as *Eomes* and *Bmp4*; blue shows *Nodal* expression; red shows expression of genes that indicate prospective anterior identity; white arrows indicate direction of cell movement. **d**, A magnified schematic diagram of the signalling pathway that is triggered in the visceral endoderm by *Nodal* (n) signalling in the epiblast (black box in **a**). S2, *Smad2*; S4, *Smad4*.

sufficient to prevent spreading of the proximal *Nodal*^{lacZ} expression domain into either the distal epiblast or visceral endoderm in both wild-type and mutant embryos (Fig. 3g and h). Finally, *Nodal* is activated in the epiblast but not the visceral endoderm of *FAST* mutant embryos (P. Hoodless and J. Wrana, personal communication) and, as predicted, we found that the intronic enhancer failed to be expressed in *FAST*-deficient embryos (J.B., D.P.N. and E. J. R., unpublished observations). Thus activity of the intronic enhancer is strictly dependent on *FAST*.

We introduced the intronic enhancer transgene and the *Nodal*^{Δ100lacZ} deletion allele into a *Smad2*-deficient background¹¹. Transcription is independent of *Smad2* in the epiblast (Fig. 3c). Similarly, both the *FAST*-dependent enhancer and the *Nodal*^{Δ100lacZ} allele are activated in the epiblast of *Smad2*-deficient embryos (Fig. 3f, i). In contrast, *Smad2* is required for *Nodal* expression in the visceral endoderm (Fig. 3c', f', i'). Collectively these data confirm that *Nodal* expression in the visceral endoderm is strictly dependent on *Smad2*, a finding consistent with the observation that *Smad2* but not the closely related family-member *Smad3* is expressed in this tissue²⁸. Moreover, recent evidence indicates that both *Smad2* and *Smad3* are activated in response to Nodal signalling¹⁰. Thus, *Smad2* and *Smad3* may function interchangeably in conjunction with *FAST* to upregulate *Nodal* expression in the epiblast.

We propose that the instructive cues responsible for patterning the proximal–distal axis of the early mouse embryo initially arise in the epiblast (Fig. 4). *Nodal* expression is induced in the proximal region, perhaps in response to localized signals from the ExE, and then spreads rapidly throughout the epiblast through a *Smad2*-independent autoregulatory mechanism. Nodal signals within the epiblast promote the expression of posterior genes, such as *Wnt3* and *Eomes*, that are required for mesoderm formation. Nodal also acts to maintain expression of key patterning molecules such as *Bmp4* and *Eomes* in the ExE. Both of these functions are also independent of *Smad2*. Finally, Nodal activity in the epiblast activates the *Smad2* pathway in the overlying visceral endoderm to establish molecular patterning and to promote formation of the AVE. In response to Nodal signals, nuclear *Smad2* complexes associate with diverse cofactors to selectively upregulate or repress target gene expression (Fig. 4d). We propose that, in addition to *Nodal* itself, critical targets are likely to include *Lhx1* and *Foxa2*, as embryos lacking both these molecules phenocopy *Smad2* mutants^{11,18}. Embryos deficient in *Nodal* and *Smad2* fail to express the earliest markers of visceral endoderm, as well as the secreted antagonists encoded by *Cer1* and *Lefty1*. Consequently, the ability to maintain expression of anterior markers such as *Otx2* in the adjacent epiblast is lost.

In sum, our experiments demonstrate that Nodal signals from the epiblast dictate molecular identity within all three tissues along the proximal–distal axis. It remains to be determined how the directed anterior movement of distal visceral endoderm cells and accompanying posteriorward movement of proximal epiblast cells ultimately gives rise to the definitive embryonic anterior–posterior axis. Recent data indicate that *Cripto* and *Otx2* are involved, as loss of either molecule causes the AVE to remain distal^{3,5,29}. Interestingly, *Cripto* has been shown to act as an essential cofactor for Nodal signalling in zebrafish³⁰. However, *Cripto* mutant embryos express posterior markers such as *T* and *Fgf8* and form extra-embryonic mesoderm³. Thus, Nodal can act through *Cripto*-independent pathways to promote posterior fates, but *Cripto* is essential for Nodal functions in the epiblast leading to displacement of the AVE. We suggest that *Cripto* may modulate the strength or possibly the duration of Nodal signalling and thus contribute to the complex regulatory mechanisms that control Nodal activity in discrete tissue sites of the early embryo. Here we show that highly dynamic patterns of *Nodal* expression in the early embryo are governed by the interplay of two distinct *cis*-regulatory regions. The intronic

enhancer functions in a FAST/Smad-dependent fashion. Much less is known about *cis*-regulatory elements lying within the 5' enhancer. Additional work is needed to provide clearer insight into spatially and temporally restricted signalling cues that act upstream to activate *Nodal* in the epiblast. □

Received 24 November 2000; accepted 30 April 2001.

1. Beddington, R. S. P. & Robertson, E. J. Axis development and early asymmetry in mammals. *Cell* **96**, 195–209 (1999).
2. Thomas, P. & Beddington, R. S. P. Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr. Biol.* **6**, 1487–1496 (1996).
3. Ding, J. *et al.* *Cripto* is required for correct orientation of the anterior–posterior axis in the mouse embryo. *Nature* **395**, 702–707 (1998).
4. Rhinn, M. *et al.* Sequential roles for *Otx2* in visceral endoderm and neuroectoderm for forebrain and midbrain induction and specification. *Development* **125**, 845–856 (1998).
5. Kimura, C. *et al.* Visceral endoderm mediates forebrain development by suppressing posteriorizing signals. *Dev. Biol.* **225**, 304–321 (2000).
6. Schier, A. F. & Shen, M. M. *Nodal* signalling in vertebrate development. *Nature* **403**, 385–389 (2000).
7. Varlet, I., Collignon, J. & Robertson, E. J. *Nodal* expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. *Development* **124**, 1033–1044 (1997).
8. Zhou, X., Sasaki, H., Lowe, L., Hogan, B. L. & Kuehn, M. R. *Nodal* is a novel TGF- β -like gene expressed in the mouse node during gastrulation. *Nature* **361**, 543–547 (1993).
9. Conlon, F. L. *et al.* A primary requirement for *Nodal* in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919–1928 (1994).
10. Kumar, A. *et al.* *Nodal* signaling uses activin and transforming growth factor- β receptor-regulated Smads. *J. Biol. Chem.* **276**, 656–661 (2000).
11. Waldrip, W. R., Bikoff, E. K., Hoodless, P. A., Wrana, J. L. & Robertson, E. J. Smad2 signaling in extra-embryonic tissues determines anterior–posterior polarity of the early mouse embryo. *Cell* **92**, 797–808 (1998).
12. Liu, P. *et al.* Requirement for *Wnt3* in vertebrate axis formation. *Nature Genet.* **22**, 361–365 (1999).
13. Russ, A. P. *et al.* *Eomesodermin* is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95–99 (2000).
14. Meno, C. *et al.* Mouse *Lefty2* and zebrafish *antivin* are feedback inhibitors of *Nodal* signaling during vertebrate gastrulation. *Mol. Cell* **4**, 287–298 (1999).
15. Rosner, M. H. *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686–692 (1990).
16. Poirier, F., Timmons, P. M., Chan, C. T., Guenet, J. L. & Rigby, P. W. Expression of the *L14* lectin during mouse embryogenesis suggests multiple roles during pre- and post-implantation development. *Development* **115**, 143–155 (1992).
17. Thomas, P. Q., Brown, A. & Beddington, R. S. P. *Hex*: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125**, 85–94 (1998).
18. Perea-Gomez, A., Shawlot, W., Sasaki, H., Behringer, R. R. & Ang, S. HNF3 β and *Lim1* interact in the visceral endoderm to regulate primitive streak formation and anterior–posterior polarity in the mouse embryo. *Development* **126**, 4499–4511 (1999).
19. Belo, J. A. *et al.* *Cerberus*-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mech. Dev.* **68**, 45–57 (1997).
20. Ang, S. L., Conlon, R. A., Jin, O. & Rossant, J. Positive and negative signals from mesoderm regulate the expression of mouse *Otx2* in ectoderm explants. *Development* **120**, 2979–2989 (1994).
21. Poirier, F. *et al.* The murine *H19* gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. *Development* **113**, 1105–1114 (1991).
22. Blum, M. *et al.* Gastrulation in the mouse: the role of the homeobox gene *goosecoid*. *Cell* **69**, 1097–1106 (1992).
23. Adachi, H. *et al.* Determination of left/right asymmetric expression of *Nodal* by a left side-specific enhancer with sequence similarity to a lefty-2 enhancer. *Genes Dev.* **13**, 1589–1600 (1999).
24. Norris, D. P. & Robertson, E. J. Asymmetric and node-specific *Nodal* expression patterns are controlled by two distinct *cis*-acting regulatory elements. *Genes Dev.* **13**, 1575–1588 (1999).
25. Whitman, M. Smads and early developmental signaling by the TGF β superfamily. *Genes Dev.* **12**, 2445–2462 (1998).
26. Saijoh, Y. *et al.* Left–right asymmetric expression of *Lefty2* and *Nodal* is induced by a signaling pathway that includes the transcription factor *FAST2*. *Mol. Cell* **5**, 35–47 (2000).
27. Collignon, J., Varlet, I. & Robertson, E. J. Relationship between asymmetric *Nodal* expression and the direction of embryonic turning. *Nature* **381**, 155–158 (1996).
28. Tremblay, K. D., Hoodless, P. A., Bikoff, E. K. & Robertson, E. J. Formation of the definitive endoderm in mouse is a Smad2-dependent process. *Development* **127**, 3079–3090 (2000).
29. Acampora, D. *et al.* Visceral endoderm-restricted translation of *Otx1* mediates recovery of *Otx2* requirements for specification of anterior neural plate and normal gastrulation. *Development* **125**, 5091–5104 (1998).
30. Gritsman, K. *et al.* The EGF-CFC protein one-eyed pinhead is essential for *Nodal* signaling. *Cell* **97**, 121–132 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. Bikoff, P. Hoodless, R. Dunn and K. Tremblay for helpful discussions and comments on the manuscript; J. Wrana and P. Hoodless for permission to cite unpublished work; and P. Lewko and J. Rocca for animal care. This work was funded by the NIH and supported by a postdoctoral fellowship from the Wellcome Trust (J.B.) and a

Charles A. King Trust fellowship (D.P.N.). This paper is dedicated to the memory of Rosa Beddington, our dear friend and colleague, who died on 18 May 2001.

Correspondence and requests for materials should be addressed to E.J.R. (e-mail: ejrobert@fas.harvard.edu).

ATR/ATM-mediated phosphorylation of human Rad17 is required for genotoxic stress responses

Shideng Bao^{*}, Randal S. Tibbetts^{*}, Kathryn M. Brumbaugh, Yanan Fang, D. Ashley Richardson, Ambereen Ali, Susan M. Chen, Robert T. Abraham & Xiao-Fan Wang

Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

^{*} These authors contributed equally to this work.

Genotoxic stress triggers the activation of checkpoints that delay cell-cycle progression to allow for DNA repair¹. Studies in fission yeast implicate members of the Rad family of checkpoint proteins, which includes Rad17, Rad1, Rad9 and Hus1, as key early-response elements during the activation of both the DNA damage and replication checkpoints^{2–5}. Here we demonstrate a direct regulatory linkage between the human Rad17 homologue (hRad17) and the checkpoint kinases, ATM and ATR. Treatment of human cells with genotoxic agents induced ATM/ATR-dependent phosphorylation of hRad17 at Ser 635 and Ser 645. Overexpression of a hRad17 mutant (hRad17^{AA}) bearing Ala substitutions at both phosphorylation sites abrogated the DNA-damage-induced G₂ checkpoint, and sensitized human fibroblasts to genotoxic stress. In contrast to wild-type hRad17, the hRad17^{AA} mutant showed no ionizing-radiation-inducible association with hRad1, a component of the hRad1–hRad9–hHus1 checkpoint complex. These findings demonstrate that ATR/ATM-dependent phosphorylation of hRad17 is a critical early event during checkpoint signalling in DNA-damaged cells.

The human checkpoint protein hRad17 and its fission (*Schizosaccharomyces pombe*, Rad17) and budding (*Saccharomyces cerevisiae*, Rad24) yeast homologues are proteins related to replication factor C (RFC) containing Walker A-type and B-type nucleotide-binding motifs^{2,6–8}. On the basis of homology to RFC, it has been proposed that Rad17 loads the Rad1–Rad9–Hus1 checkpoint complex onto damaged DNA at an early stage of activation of the cell-cycle checkpoint^{6,9,10}. To further define the genome-surveillance functions of hRad17, we searched for interactions between hRad17 and other known checkpoint signalling proteins. In immunoprecipitation experiments, we found that antibodies directed against the two related checkpoint kinases, ATR and ATM¹¹, coprecipitated hRad17 from MCF-7 breast carcinoma cell extracts. The amount of ATR-associated hRad17 was strongly increased after cellular exposure to ultraviolet light, ionizing radiation or to the DNA-replication inhibitor HU (Fig. 1a, top panel). In parallel experiments, ATM also exhibited an interaction with hRad17 that was inducible by DNA damage; however, in contrast to the results obtained with ATR, the association of ATM with hRad17 was selectively induced by ionizing radiation (Fig. 1b, top panel).

The inducible association between ATR/ATM and hRad17 indicated that hRad17 might be a substrate for these checkpoint kinases. Comparison of the primary sequences of human, mouse and monkey Rad17 revealed two conserved Ser–Gln motifs at Ser 635 and Ser 645 in hRad17. The presence of Gln at the +1 position

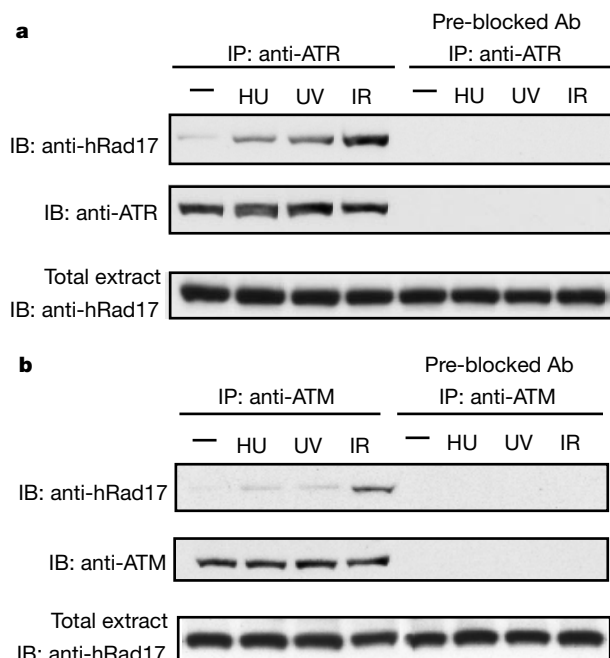


Figure 1 Association of hRad17 with ATR/ATM induced by DNA damage. **a**, Association of hRad17 with ATR induced by HU, ultraviolet light (UV) and ionizing radiation (IR). MCF-7 cells were either left untreated, cultured for 2 h in 10 mM HU, or exposed to 50 J m⁻² ultraviolet light or 25 Gy ionizing radiation, and collected 1 h later. Cellular extracts were immunoprecipitated (IP) with anti-ATR antibodies, or the same antibodies (Ab) pre-blocked with immunizing peptide. Immunoprecipitated proteins were immunoblotted (IB) with anti-hRad17 or anti-ATR antibodies. **b**, Association of hRad17 with ATM induced by ionizing radiation. MCF-7 cells were treated as described in **a**. Immunoprecipitations were performed with anti-ATM antibodies, and proteins were immunoblotted with anti-hRad17 or anti-ATM antibodies. Total expression of hRad17 was determined with one-fifteenth of the volume of the total cell extracts used in **a** and **b**.

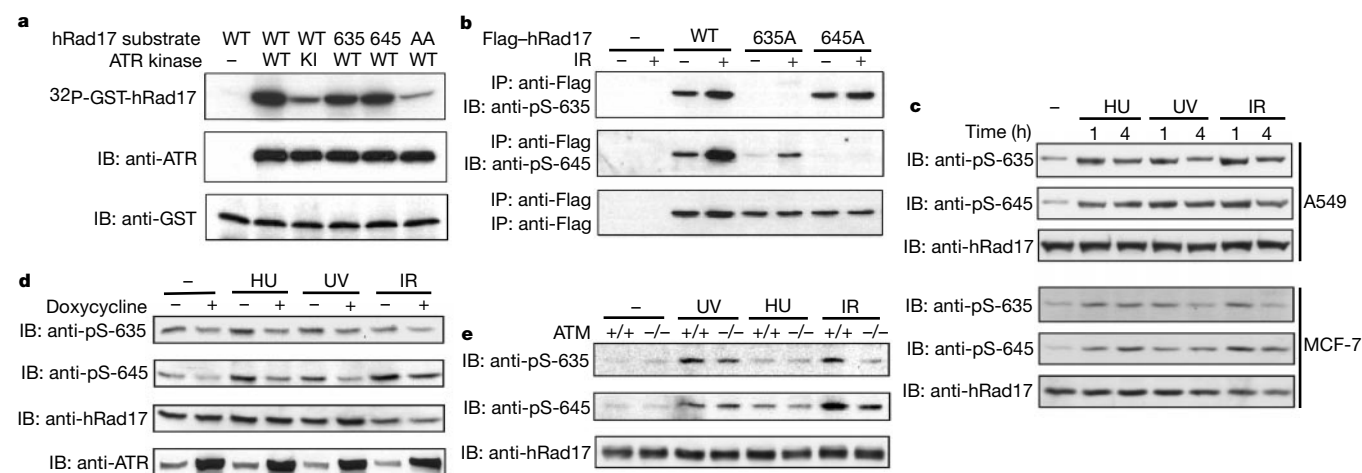


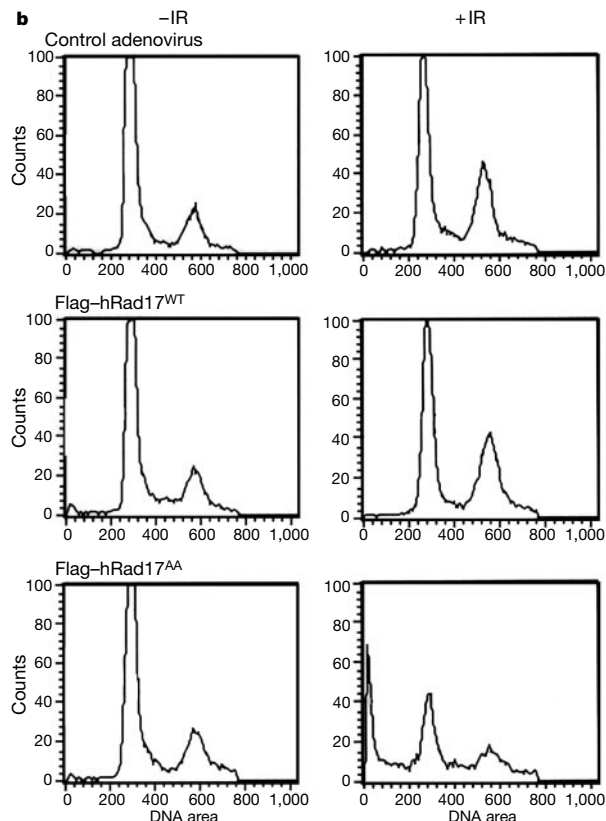
Figure 2 Phosphorylation of hRad17 induced by DNA damage. **a**, ATR phosphorylates hRad17 *in vitro*. HEK 293T cells were transfected with pcDNA3 only (—), or with Flag-tagged wild-type (WT) or kinase-inactive (KI) ATR-expression plasmids. Cellular extracts were immunoprecipitated with anti-Flag monoclonal antibody, and kinase assays were performed with wild-type or mutated GST-hRad17 substrates containing single (635, 645) or double (AA) Ala substitutions at Ser 635 and Ser 645. ³²P-labelled reaction products are shown. Amounts of ATR and GST-hRad17 were determined by immunoblotting (IB). **b**, hRad17 phosphorylation in intact cells. Cells (293T) were transfected with expression plasmids encoding wild-type (WT) or mutated (635A or 645A) Flag-hRad17. Cells were left untreated (—) or exposed to 20 Gy ionizing radiation (+), and collected 1 h after irradiation. Anti-Flag immunoprecipitates (IP) were immunoblotted with

suggested that either or both of these residues might be preferred target sites for phosphorylation by ATR or ATM^{12–14}. To determine whether ATR phosphorylates hRad17, we immunoprecipitated Flag-tagged ATR from human embryonic kidney (HEK) 293T cells and performed immune-complex kinase assays with a glutathione S-transferase (GST)-hRad17 (residues 486–670) fusion protein as the substrate. The results demonstrated that wild-type ATR phosphorylated GST-hRad17 at a level significantly above the background observed with the kinase-inactive ATR^{KI} (Fig. 2a, top panel). Phosphorylation of the GST-hRad17 substrate was reduced by ~50% when either Ser 635 or Ser 645 was replaced with Ala, and a GST-hRad17 double mutant containing Ala substitutions at both sites reduced ATR-mediated phosphorylation to near-background levels. These findings demonstrate that ATR phosphorylates the hRad17 carboxy-terminal region exclusively at Ser 635 and Ser 645, and are consistent with previous results obtained with hRad17-derived peptides as substrates¹².

To determine whether hRad17 is phosphorylated on Ser 635 or Ser 645 in intact cells, we generated phosphospecific antibodies (anti-pS-635 and anti-pS-645) directed against each phosphorylation site. The specificities of these antibodies were first tested in HEK 293T cells expressing various Flag-tagged hRad17 constructs. Both anti-pS-635 and anti-pS-645 antibodies recognized the wild-type Flag-hRad17 protein in immunoblot analyses (Fig. 2b). Notably, cellular treatment with ionizing radiation enhanced the recognition of Flag-hRad17 by both phosphospecific antibodies, which suggested that phosphorylation of Ser 635 and Ser 645 was stimulated by ionizing radiation. Mutation of either Ser 635 or Ser 645 to an Ala residue that could not be phosphorylated eliminated the recognition of hRad17 by the corresponding phosphospecific antibody (Fig. 2b). Furthermore, pretreatment of the immunoprecipitated Flag-hRad17 with λ-phosphatase abolished the detection of this protein by anti-pS-635 and anti-pS-645 in immunoblot analyses (data not shown). Collectively, these results document the specificities of the anti-pS-635 and anti-pS-645 antibodies, and indicate that hRad17 undergoes increased phos-

phosphorylation. **c**, Phosphorylation of hRad17 induced by DNA damage. A549 cells and MCF-7 cells were either left untreated (—) or exposed to HU, ultraviolet light (UV) or ionizing radiation (IR). After 1 or 4 h, cellular extracts were immunoblotted with the indicated antibodies. **d**, Overexpression of kinase-inactive ATR inhibits hRad17 phosphorylation. GM847/ATR^{KI} cells were cultured for 48 h in the absence (—) or presence (+) of doxycycline to induce ATR^{KI} expression. The cells were either left untreated or exposed to 10 mM HU, 50 J m⁻² ultraviolet light or 20 Gy ionizing radiation. After 1 h, cellular extracts were analysed by immunoblotting. **e**, Phosphorylation of hRad17 in ATM-deficient cells. ATM^{+/+} or ATM^{-/-} mouse embryonic fibroblasts were treated with the indicated agents, and analysed as in **d**.

We next determined whether functional ATR or ATM is required for genotoxic-stress-induced phosphorylation of hRad17 on Ser 635 and Ser 645. As a cellular model for ATR deficiency, we used a GM847 human fibroblast cell line that overexpresses a Flag-tagged, kinase-inactive ATR mutant (ATR^{KI}) when cultured in the presence of doxycycline^{13,15}. In the absence of doxycycline, treatment of these cells with HU or ultraviolet light induced the phosphorylation of hRad17 on Ser 635 and Ser 645 (Fig. 2d). Overexpression of Flag-ATR^{KI} suppressed the phosphorylation of both sites in GM847 cells treated with HU or ultraviolet light (Fig. 2d). The phosphorylation of Ser 635 and Ser 645 in undamaged cells was also significantly reduced by overexpression of ATR^{KI}, suggesting that ATR contributes to the basal phosphorylation of hRad17 in cycling cells.



a

Flag-hRad17: WT AA

IR: - + - +

IP: anti-ATR

IB: anti-Flag

IP: anti-ATM

IB: anti-Flag

Total extract

IB: anti-Flag

b

CV WT AA

IR: - + - + - +

IB: anti-pS-635

IB: anti-hRad17

IB: anti-Flag

Flag-hRad17

Endogenous hRad17

Flag-hRad17

Endogenous hRad17

Flag-hRad17

Figure 4 Overexpression of Flag-hRad17^{AA} suppresses phosphorylation of hRad17. **a**, Binding of the Flag-hRad17^{AA} (AA) protein to ATR and ATM induced by ionizing radiation (IR). Co-immunoprecipitation (IP) assays were performed as described in Fig. 1. WT, wild-type Flag-hRad17^{WT}; IB, immunoblotting. **b**, Adenovirus-infected BJ cells (CV, control adenovirus; WT, Flag-hRad17^{WT}; AA, Flag-hRad17^{AA}) were either left untreated or exposed to 20 Gy ionizing radiation, and collected 1 h after irradiation. Cellular extracts were analysed by sequential immunoblotting with the indicated antibodies.

20 h, cell-cycle distributions were determined by flow cytometry. **c**, Adenovirus-infected BJ cells (CV, control adenovirus; WT, Flag-hRad17^{WT}; AA, Flag-hRad17^{AA}) were exposed to ultraviolet light (UV) or HU, and percentages of cells with hypodiploid DNA content were quantified by flow cytometry. **d**, Overexpression of Flag-hRad17^{AA} overrides checkpoint-mediated G₂ arrest. Adenovirus-infected BJ cells were irradiated with 20 Gy ionizing radiation. Mitotic cells were counted as described in the Methods.

Interestingly, Ser 645 phosphorylation induced by ionizing radiation was more resistant to ATR^{K1} overexpression than was the response to either HU or ultraviolet light, which is consistent with the notion that ATM is the major effector of Ser 645 phosphorylation in cells damaged by ionizing radiation.

Previous results suggest that ATM is important in cell-cycle-checkpoint activation triggered by ionizing radiation and radiomimetic agents^{11,13,16}. Consequently, we compared the phosphorylation of Ser 635 and Ser 645 in ATM^{+/+} and ATM^{-/-} mouse embryonic fibroblasts. Immunoblot analysis revealed that, while exposure to HU or ultraviolet light induced similar levels of Ser 635 and Ser 645 phosphorylation in ATM^{+/+} and ATM^{-/-} cells, phosphorylation stimulated by ionizing radiation at both sites was clearly compromised in the ATM^{-/-} cells (Fig. 2e). Collectively, these results indicate that the phosphorylation of hRad17 in cells damaged by ionizing radiation is heavily dependent upon the expression of ATM, whereas the response to other forms of genotoxic stress is mediated primarily through ATR.

The functional consequences of ATR/ATM-dependent phosphorylation of hRad17 were examined through adenovirus-mediated overexpression of a Flag-hRad17 double mutant containing Ala substitutions at Ser 635 and Ser 645 (Flag-hRad17^{AA}) in human diploid BJ fibroblasts. In preliminary studies, we confirmed that these cells expressed comparable levels of Flag-hRad17^{WT} (wild type) or Flag-hRad17^{AA} protein 24 h after infection with recombinant adenovirus (Fig. 3a). To determine whether overexpression of Flag-hRad17^{AA} affected cell-cycle arrest or cell survival following DNA damage, we infected BJ fibroblasts with control adenovirus or recombinant adenoviruses encoding Flag-hRad17^{WT} or Flag-hRad17^{AA}. The virus-infected cells were exposed to ionizing radiation, and cell-cycle distributions were analysed 20 h after irradiation. BJ cells that were infected with either control or Flag-hRad17^{WT}-encoding adenoviruses displayed a pronounced G₂/M

arrest after exposure to ionizing radiation (Fig. 3b). In contrast, the Flag-hRad17^{AA}-expressing cells failed to accumulate in G₂/M phase after exposure to ionizing radiation. Instead, many of the Flag-hRad17^{AA}-expressing cells underwent apoptotic cell death, as indicated by the dramatic increase in the percentage of cells with hypodiploid DNA contents, from 2% in the control population to 34% in the cells treated with ionizing radiation (Fig. 3b, bottom panels). Overexpression of Flag-hRad17^{AA}, but not Flag-hRad17^{WT}, caused similar increases in cellular sensitivity to ultraviolet light or HU exposure (Fig. 3c).

The studies of cell-cycle distribution described above indicated that overexpression of the Flag-hRad17^{AA} mutant may override the G₂ checkpoint. We examined this possibility through measurement of the mitotic indices of γ -irradiated BJ fibroblasts infected with Flag-hRad17^{WT}- or Flag-hRad17^{AA}-encoding adenoviruses. Nocodazole was added 1 h after irradiation, to capture in mitosis all cells that had bypassed the G₂ checkpoint. BJ fibroblasts infected with either the control or Flag-hRad17^{WT}-encoding adenovirus displayed a nearly complete block to mitotic entry following exposure to ionizing radiation (Fig. 3d). In contrast, the cells overexpressing Flag-hRad17^{AA} displayed no reduction in mitotic index after treatment with ionizing radiation, indicating that phosphorylation of hRad17 on Ser 635 and Ser 645 was required for activation of the G₂ checkpoint by ionizing radiation (Fig. 3d).

To gain further insights into the mechanism underlying the dominant suppressive effect of Flag-hRad17^{AA} on checkpoint responses to DNA damage, we first examined whether the Flag-hRad17^{AA} mutant retained the ability to interact with ATR/ATM after treatment with ionizing radiation. The Ser-to-Ala substitutions had no effect on the ionizing-radiation-induced association of Flag-hRad17^{AA} with either ATR or ATM (Fig. 4a). However, consistent with its ability to interact with ATM/ATR, the Flag-hRad17^{AA} mutant strongly reduced the basal phosphorylation of

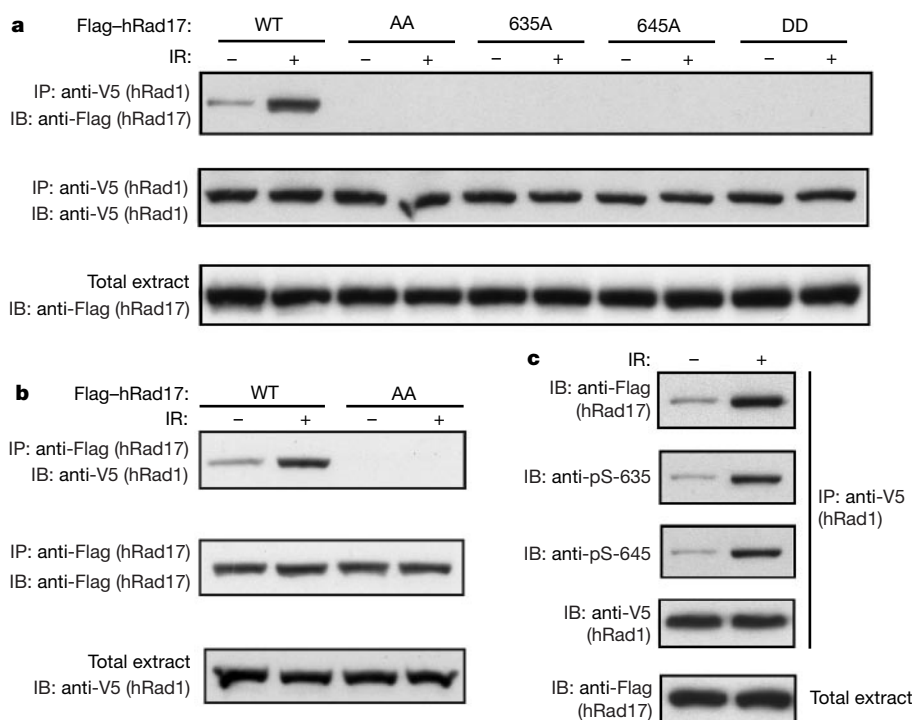


Figure 5 Association of hRad17 with hRad1 induced by DNA damage. **a**, HEK 293 cells were co-transfected with plasmids encoding V5-tagged hRad1 and the indicated Flag-hRad17 proteins. The transfected cells were either left untreated or exposed to 20 Gy ionizing radiation, and collected after 1 h. Cellular extracts were immunoprecipitated (IP) with anti-V5, and analysed by immunoblotting (IB) with anti-Flag or anti-V5 as indicated.

b, Samples were prepared as described in **a**, except that co-precipitation was performed in a reciprocal fashion. **c**, HEK 293 cells were co-transfected with Flag-hRad17^{WT} and V5-hRad1 expression plasmids. The anti-V5 immunoprecipitates were immunoblotted with anti-pS-635 and anti-pS-645 antibodies.

endogenous hRad17 on Ser 635, and blocked the ionizing-radiation-induced phosphorylation at this site in BJ fibroblasts (Fig. 4b). These findings show that the inhibitory effects of Flag-hRad17^{AA} on checkpoint signalling are explained, at least in part, through disruption of the ATM/ATR-dependent pathway leading to the phosphorylation of endogenous hRad17 in these cells.

In subsequent studies, we determined whether phosphorylation of hRad17 on Ser 635 and Ser 645 modulates the proposed interaction between hRad17 and the hRad1-hRad9-hHus1 checkpoint complex. The hRad1-hRad9-hHus1 complex seems to form a checkpoint-sliding clamp (CSC), similar to the proliferating-cell nuclear antigen (PCNA) homotrimer, which functions in the processing of damaged DNA and in the initiation of checkpoint signalling^{2,6,9,10,17}. According to this model, hRad17 interacts with other RFC subunits to facilitate loading of the CSC onto damaged DNA^{6,10,17}. To determine whether phosphorylation on Ser 635 and Ser 645 regulates interactions between hRad17 and components of the CSC, we performed co-immunoprecipitation experiments in HEK 293 cells that were co-transfected with V5-epitope-tagged hRad1 and either Flag-hRad17^{WT} or Flag-hRad17^{AA}. Treatment of the cells with ionizing radiation induced the appearance of Flag-hRad17^{WT}, but not Flag-hRad17^{AA}, in the anti-V5-hRad1 immunoprecipitates (Fig. 5a). Identical results were obtained in reciprocal co-precipitation experiments (Fig. 5b). To determine the contributions of the individual phosphorylation sites to the DNA-damage-induced interaction with hRad1, similar studies were performed with Flag-hRad17 single mutants containing Ala substitutions at either Ser 635 or Ser 645 (Fig. 5a). Neither of the Flag-hRad17 single mutants co-precipitated with V5-hRad1 after cellular exposure to ionizing radiation, suggesting that phosphorylation at both Ser 635 and Ser 645 was required for the inducible interaction of hRad17 with hRad1. Immunoblot analyses with the phosphospecific anti-pS-635 and anti-pS-645 antibodies confirmed that the V5-hRad1-associated pool of Flag-hRad17 was phosphorylated on both sites (Fig. 5c). Although the functional consequences of Ser or Thr phosphorylation can sometimes be duplicated by substitutions of the targeted residues with acidic amino acids, we observed that a Flag-hRad17^{DD} mutant failed to associate with V5-hRad1 in cells damaged by ionizing radiation (Fig. 5a). Thus, phosphorylation of hRad17 by ATM/ATR modifies the protein in a fashion that cannot be mimicked by the presence of negatively charged residues alone.

This report demonstrates that genotoxic stress triggers the association of ATM and ATR with hRad17 and the resultant phosphorylation of hRad17 at sites required for G₂ checkpoint function in mammalian cells. Accumulating evidence implicates hRad17 as a key component of the PCNA-like checkpoint-loading complex, which loads the hRad1-hRad9-hHus1 CSC at or near sites of DNA damage². Our results further indicate that ATM/ATR-dependent hRad17 phosphorylation may regulate the interaction between hRad17 and the CSC in DNA-damaged cells. The conclusion that ATM and ATR are upstream regulators of hRad17 function reinforces the idea that these protein kinases serve as very early signalling elements in checkpoint pathways induced by DNA damage. □

Methods

Cell culture and antibodies

The MCF-7, A549, HEK 293 and HEK 293T cell lines were cultured in DMEM with 10% FBS. GM847 fibroblast cells were cultured as described¹⁵. Human BJ fibroblasts and ATM^{-/-} and ATM^{+/+} mouse embryo fibroblasts were maintained in DMEM containing 20% FBS. Where indicated, cells were irradiated with a ¹³⁷Cs gamma ray source or exposed to a ultraviolet-C light source of wavelength 254 nm. Phosphospecific antibodies directed against hRad17 were generated by immunizing rabbits with keyhole-limpet haemocyanin-conjugated phosphopeptides spanning the Ser 635 (CETWSLPSPQNSASE) and Ser 645 (CSASELPAPSPQPFPS) phosphorylation sites. The phosphospecific antibodies were affinity purified from rabbit antisera as previously described¹⁸. The anti-ATR (C-19 and Ab-1) and anti-ATM (K-19, H-248 and Ab-3) antibodies, and their corresponding

blocking peptides, were from Santa Cruz and Oncogene Research Products. Anti-Flag (M2) monoclonal antibody and the anti-Flag M2 agarose affinity gel were from Sigma. The anti-V5 antibody and anti-V5 conjugated to horseradish peroxidase were obtained from Invitrogen.

Protein analysis and kinase assays

To examine the association between hRad17 and ATR/ATM, cells were extracted with lysis buffer (20 mM HEPES at pH 7.4, 100 mM NaCl, 40 mM KCl, 1.5 mM MgCl₂, 1 mM EGTA and 0.5% NP-40) supplemented with phosphatase inhibitors (10 mM NaF, 1 mM Na₃VO₄, 50 mM β-glycerophosphate and 25 nM microcystin-LR) and protease inhibitors (20 μg ml⁻¹ leupeptin, 10 μg ml⁻¹ pepstatin A and 10 μg ml⁻¹ aprotinin). The cleared extracts were immunoprecipitated with 4 μg of anti-ATR (C-19) or anti-ATM (K-19) antibodies for 2 h at 4 °C. The immunoprecipitates were washed four times with lysis buffer, solubilized with SDS-PAGE sample buffer, and electrophoresed as described¹³. Proteins were transferred to Immobilon P membranes (Millipore) and then immunoblotted with anti-hRad17, anti-ATR (Ab-1), anti-ATM (Ab-3), anti-pS-635 or anti-pS-645 (1 μg ml⁻¹ for each antibody).

The association of Flag-hRad17 with V5-hRad1 was evaluated by co-transfection of HEK 293 cells with pcDNA3-V5-hRad1 and either pCMV2-Flag-hRad17^{WT}, pCMV2-Flag-hRad17^{AA}, pCMV2-Flag-hRad17^{635A}, pCMV2-Flag-hRad17^{645A} or pCMV2-Flag-hRad17^{DD}. At 24 h after transfection, the cells were treated with 20 Gy ionizing radiation, and then lysed after 1 h in HS buffer (20 mM HEPES at pH 7.4, 100 mM KCl, 60 mM NaCl, 10 mM MgCl₂, 4 mM MnCl₂, 10 mM NaF and 70 μg ml⁻¹ digitonin, supplemented with 10 mM β-glycerophosphate, 1 mM Na₃VO₄, 25 nM microcystin-LR and protease inhibitors). The cleared extracts were diluted with an equal volume of water and then immunoprecipitated for 2 h at 4 °C with anti-Flag or anti-V5 monoclonal antibody and protein G-Sepharose. The immunoprecipitates were washed four times with a 1:2 dilution of HS buffer, and analysed by immunoblotting with anti-V5-HRP or anti-Flag. ATR kinase assays were as previously described¹³ with 1 μg per reaction of GST-hRad17 fusion protein containing the C-terminal 185 amino acids of hRad17 as the substrate.

Plasmid and adenoviral constructs

The pFLAG-hRad17/CMV2 vector was constructed by cloning the hRad17 complementary DNA into pFLAG-CMV2 using the *Clal*-*Bam*H1 sites. Flag-hRad17 expression vectors containing Ser→Ala or Ser→Asp substitutions at Ser 635 or Ser 645 were generated with the Quick Change mutagenesis kit (Stratagene). The V5-hRad1/pcDNA3 plasmid was constructed by cloning a V5-tagged hRad1 cDNA construct into pcDNA3 using *Eco*R1-*Xho*I sites. GST-hRad17 was constructed by subcloning a hRad17 cDNA fragment encoding amino acids 486-670, amplified by polymerase chain reaction, into the *Bam*HI site of pGEX-2T. The generation and production of recombinant adenoviruses were carried out as described¹⁹. Briefly, Flag-hRad17^{WT} and Flag-hRad17^{AA} were cloned into the pAdTrack-CMV shuttle vector at *Kpn*I-*Xho*I sites, and then co-transformed with the adenoviral backbone vector pAdEasy-1 into *Escherichia coli* BJ5183 cells to generate recombinant adenoviral plasmids. These plasmids were transfected into HEK 293 cells for the production of high-titre adenoviral stocks.

FACS analysis and mitotic indices

Human BJ fibroblasts were infected for 24 h with recombinant adenoviruses. Infected cells were irradiated with 20 J m⁻² ultraviolet light or 15 Gy ionizing radiation, or were cultured in medium containing 10 mM HU. After 20 h, cells were fixed with 70% ethanol and incubated for 30 min with RNase A (100 μg ml⁻¹) and propidium iodide (50 μg ml⁻¹) at 37 °C. Cell-cycle distributions were analysed by flow cytometry. To determine mitotic indices, BJ fibroblasts were cultured overnight on coverslips, and then infected with recombinant adenoviruses. After 24 h, the infected cells were either left untreated or treated with ionizing radiation (20 Gy). One hour after irradiation, the cells were treated with nocodazole (0.5 μg ml⁻¹) and cultured for an additional 12 h. Cells were then fixed with 4% paraformaldehyde and stained with DAPI. The mitotic and interphase cells were visualized and counted by immunofluorescence microscopy.

Received 31 January; accepted 23 May 2001.

1. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433-439 (2000).
2. O'Connell, M. J., Walworth, N. C. & Carr, A. M. The G₂-phase DNA-damage checkpoint. *Trends Cell Biol.* **10**, 296-303 (2000).
3. Lowndes, N. F. & Murguia, J. R. Sensing and responding to DNA damage. *Curr. Opin. Gen. Dev.* **10**, 17-25 (2000).
4. Weinert, T. Yeast checkpoint controls and relevance to cancer. *Cancer Surveys* **29**, 109-132 (1997).
5. Caspari, T. & Carr, A. M. DNA structure checkpoint pathways in *Schizosaccharomyces pombe*. *Biochimie* **81**, 173-181 (1999).
6. Venclovas, C. & Thelen, M. P. Structure-based predictions of Rad1, Rad9, Hus1 and Rad17 participation in sliding clamp and clamp-loading complexes. *Nucleic Acids Res.* **28**, 2481-2493 (2000).
7. Bao, S., Shen, X., Shen, K., Liu, Y. & Wang, X. F. The mammalian Rad24 homologous to yeast *Saccharomyces cerevisiae* Rad24 and *Schizosaccharomyces pombe* Rad17 is involved in DNA damage checkpoint. *Cell Growth Diff.* **9**, 961-967 (1998).
8. Parker, A. E., Van, D. W., Laus, M. C., Verhasselt, P. & Luyten, W. H. M. L. Identification of a human homologue of the *Schizosaccharomyces pombe* rad17(+) checkpoint gene. *J. Biol. Chem.* **274**, 24438-24439 (1999).
9. Thelen, M. P., Venclovas, C. & Fidelis, K. A sliding clamp model for the Rad1 family of cell cycle checkpoint proteins. *Cell* **96**, 769-770 (1999).
10. Rauen, M., Burtelow, M. A., Dufault, V. M. & Karnitz, L. M. The human checkpoint protein hRad17 interacts with the PCNA-like proteins hRad1, hHus1, and hRad9. *J. Biol. Chem.* **275**, 29767-29771 (2000).

11. Shiloh, Y. ATM and ATR: networking cellular responses to DNA damage. *Curr. Opin. Gen. Dev.* **11**, 71–77 (2001).
12. Kim, S. T., Lim, D. S., Canman, C. E. & Kastan, M. B. Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37538–37543 (1999).
13. Tibbetts, R. S. *et al.* A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev.* **13**, 152–157 (1999).
14. Banin, S. *et al.* Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674–1677 (1998).
15. Cliby, W. A. *et al.* Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *EMBO J.* **17**, 159–169 (1998).
16. Tibbetts, R. S. *et al.* Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes Dev.* **14**, 2989–3002 (2000).
17. Burtelow, M. A., Kaufmann, S. H. & Karnitz, L. M. Retention of the human Rad9 checkpoint complex in extraction-resistant nuclear complexes after DNA damage. *J. Biol. Chem.* **275**, 26343–26348 (2000).
18. Sekulic, A. *et al.* A direct linkage between the phosphoinositide 3-kinase-AKT signaling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. *Cancer Res.* **60**,

3504–3513 (2000).

19. He, T. C. *et al.* A simplified system for generating recombinant adenoviruses. *Proc. Natl Acad. Sci. USA* **95**, 2509–2514 (1998).

Acknowledgements

We thank B. Vogelstein for the pAdTrack-CMV and pAdEasy adenoviral vectors; Y. Xu for the *ATM*^{+/+} and *ATM*^{-/-} mouse embryo fibroblasts; C. Counter for the BJ fibroblasts; L. M. Karnitz for the anti-hRad17 antibody; L. Martinek and M. Cook for the flow cytometric analysis; Y. Yu for technical help; and members of the Wang and Abraham laboratories for scientific discussions. This work was supported by grants from the National Institutes of Health, the A-T Children's Project, and the Johnson and Johnson Foundation.

Correspondence and requests for materials should be addressed to X.-F.W. (e-mail: wang0011@mc.duke.edu).

corrections

Gain-assisted superluminal light propagation

L. J. Wang, A. Kuzmich & A. Dogariu

Nature **406**, 277–279 (2000).

It has been drawn to our attention that we inadvertently did not cite a previous paper by Segard and Macke¹ on the topic of observed superluminal light pulse propagation in media with heavy absorption using a direct pulse shape detection technique. We did, however, cite an earlier paper by Chu and Wong² on this topic using a different technique. Our introductory statement “in all previous experimental demonstrations, the light pulses experienced either very large absorption or severe reshaping” stands. We have now included this reference in a more detailed description³ of our experiment. □

1. Segard, B. & Macke, B. Observation of negative velocity pulse propagation. *Phys. Lett. A* **109**, 213–216 (1985).
2. Chu, S. & Wong, S. Linear pulse propagation in an absorbing medium. *Phys. Rev. Lett.* **48**, 738–741 (1982).
3. Dogariu, A., Kuzmich, A. & Wang, L. J. Transparent anomalous dispersion and superluminal light pulse propagation at a negative group velocity. *Phys. Rev. A* **63**, 053806 (2001).

Genomic analysis of metastasis reveals an essential role for RhoC

E. A. Clark, T. R. Golub, E. S. Lander & R. O. Hynes

Nature **406**, 532–535 (2000).

A citation of previous work by del Paso *et al.*¹ as evidence in support of our results was erroneously removed during editing of an earlier draft. □

1. del Paso, L., Hernandez-Alcoceba, R., Embade, N., Carnero, A., Esteve, P., Paje, C. & Lacal, J. C. *Oncogene* **15**, 3047–3057 (1997).

Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses

R. I. Wilson & R. A. Nicoll

Nature **410**, 588–592 (2001).

It has come to our attention that we failed to cite a relevant study (Egertová, M., Giang, D. K., Cravatt, B. F. & Elphick, M. R. *Proc. R. Soc. Lond. B* **265**, 2081–2085; 1998) in our Letter. Based on the immunolocalization of fatty acid amide hydrolase, which inactivates endocannabinoids, to hippocampal pyramidal cells, and the localization of the CB1 receptor to presynaptic terminals, Egertová *et al.* proposed that endocannabinoids might function as retrograde messengers. □

Protein growth area

Cunning plans such as distributed computing are speeding up proteomics.

Omnigrid Accent

GeneMachines www.genemachines.com

Benchtop microarrayer

The OmniGrid Accent high-speed arrayer prints with precision from three plates on to 50 slides and supports a print-head with up to 48 pins. Arrays can be printed with centre-to-centre spot distances adjustable in 2.5- μ m increments. Software enables the user to control array design, pin configuration, cleaning protocols and sample sequences.

WorksBase

Bio-Rad www.proteomeworks.bio-rad.com

Data management software for Proteome Works

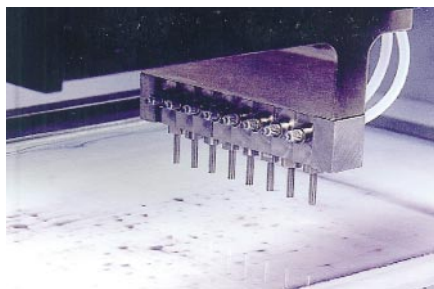
WorksBase will store, organize and analyse proteomics information, and operates in an e-MS enabled client-server environment. The program acts as a centralized warehouse in which data generated by ProteomeWorks can be integrated with external database information. ProteomeWorks includes proteomics products and analysis tools from Bio-Rad and MicroMass. The WorksBase system is designed to support most existing bioinformatics applications and includes software development kits for those wishing to write interfaces for integration of confidential information. The program allows proteomics data to be integrated with other sources of information, including genomic, chromatographic array, metabolic, biochemical and molecular modelling.

gelPix

Genetix www.genetixgroup.com

Spot on for proteomics scientists

GelPix is a high-throughput protein spot excision system that automates the processing of two-dimensional protein gels. The gels are imaged using a high-resolution camera with specific image analysis software that generates a protein target excision list. Selected proteins are then excised using an eight-channel excision head. The system can be used with standard and fluorescent stained gels.



gelPix: automation for proteomics.



The OmniGrid Accent: bench-top microarray technology.

BacterioMatch

Stratagene www.stratagene.com

Bacterial two-hybrid detection system

Described as a "fast and easy alternative to other two-hybrid systems for *in vivo* detection of protein-protein interactions", BacterioMatch carries out all screening in bacteria, with colonies growing by the next day. Methods include standard transformation and plating protocols. The system facilitates screening of larger libraries due to the high transformation efficiency of *E. coli*. Protein-protein interaction is detected as growth of ampicillin-resistant colonies. A secondary β -galactosidase marker is included to help validate putative positives.

Tp-Blue & Tp-Red

Biocolor www.biocolor.co.uk

Total protein assay kits

These enhanced dye-protein binding quantitative assays incorporate extended sensitivity, increased stability and improved safety. They are methanol-free and are based on Coomassie brilliant blue G and pyrogallol red molybdate. The assays exhibit linear calibration and samples can be read at any time between 20 minutes and 10 hours. Supplied in kits of 1,000 assays for 96-well format or 4,000 assays for 384-well format, they are suitable for cuvette and microplate analysis.

ProteinMine

Scimagix www.scimagix.com

Mine, all mine

Researchers who need to make visual comparisons of protein patterns will appreciate this software application, which is based on the

Scientific Image Management System (SIMS). ProteinMine will conduct high-throughput searches of thousands of gels to find the selected combination of protein patterns and differentially expressed proteins. Two-dimensional electrophoresis gels are searched according to visual criteria — intensity, size, shape and location of protein spots. The program can be used in conjunction with SIMS to retrieve data using a web browser, to store, analyse, search and mine information across any number of two-dimensional gels and to integrate the results with other experimental data.

ProteoPrep

Sigma-Aldrich www.sigma-aldrich.com

New line of sample preparation kits

A trio of extraction kits and two stand-alone reagents make up this line of proteomic-specific products. The universal extraction kit can be used for testing or optimizing extraction conditions to produce total protein extracts; the membrane extraction kit is for total extraction of membrane-associated proteins; and the sample extraction kit will test or optimize extraction conditions to produce total protein extracts. Reagents for reducing and alkylating complete the package. All products are supplied pre-measured and pre-mixed.

RTS 500

Roche Molecular Biochemicals

www.proteinexpression.com

Rapid translation system

Said to be superior to *E. coli* systems for purifying proteins, RTS 500 is designed to speed up the expression of target proteins for lead-generation programs. Combined with RTS

reagent sets, it provides continuous-exchange, cell-free technology for preparative-scale protein expression. Reactions can be set up in 30 minutes and can produce 500 µg in 24 hours.

Coomassie dry protein assay plates

Pierce www.piercenet.com

High-volume total protein analysis

These assay plates can be used to accelerate the performance of Bradford Coomassie dye-based assays. Each strip plate allows up to 96 individual total protein assays to be performed with no need to prepare or pipette reagent. The sample or standard is added to the microplate well, mixed, and the results read. The plates have a working range of 38–300 µg per ml and a coefficient of variation of 0.270. The sealed-strip design allows the user to select as few or as many microplate wells as needed.

2001–2003 catalogue

Peptides International www.pepnet.com

Products for peptide research

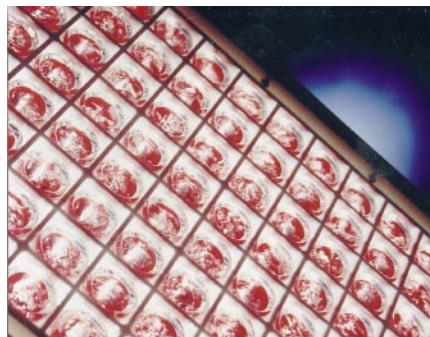
Metalloprotease inhibitor TAPI-2, calcium-channel blocker SNX-482, Ghrelin peptides, amyloid β-protein fragments, glucagon-like peptides, orexins, prolactin-releasing peptides, nocistatins and urotensins are some of the products in this new catalogue. For synthesis, there are Fmoc-CLEAR resins and orthogonal amino acid derivatives. Caspase-related peptides, peptide toxins, enzyme inhibitors endothelins, defensin peptides and MCA fluorescent substrates are also offered. The catalogue is available via the company's website.

TurboBLAST

TurboGenomics www.turbogenomics.com

Distributed computing is a BLAST

Completion of large-scale sequencing projects and the ever-growing sequence databases make great demands on BLAST processing, the sequence comparison method used to examine the structure and function of genes and proteins. TurboBlast accelerates the performance of BLAST by partitioning BLAST computations into manageable tasks and distributing



Well red: rapid protein analysis from Pierce.

them across a network of workstations. Initial trials with 11 PCs running Linux reduced a month-long BLAST run to just two days.

VistaLogic

Genometrix www.genometrix.com

Dual-function software

This software program is designed to handle the large number of data points generated by high-throughput, microarray-based genomic analyses. It can be used to analyse both gene expression and genotyping data, making it possible to integrate polymorphism discovery and association data with expression profiles and, potentially, to correlate polymorphisms with functional significance, such as transcript splicing variants or promoter regulation. The program can be used to assimilate data from multiple experiments, combining extensive sample information with array results and analysing both gene expression profiles and single nucleotide polymorphism frequencies within one application.

PCR screening kits

Advanced GenoTechs www.geno-gtac.co.jp

All in a day's work

Advanced GenoTechs' PCR 'day screening' kits are designed for rapid selection of 384-well microtitre plates in which target clones have been arrayed. DNA pools prepared from microtitre plates are dispensed into 96-PCR plates using replica-pins and dried as PCR screening kits. Kits are supplied as a custom ser-

vice for rapid and large-scale screening. Kits are prepared from RPCI-11 human BAC, human PAC and RRPCI-23 mouse BAC libraries.

Application brochure

Millipore www.millipore.com

Concentrate on proteins

This 12-page brochure is aimed at protein chemists and provides information on purification and analysis in protein research, including product specifications and literature listings to help researchers locate the most appropriate product for their application. The brochure covers preparative purification, detailed protein analysis via mass spectrometry, sequencing, NMR and other applications.

SlideHyb

Ambion www.ambion.com

Custom kit for glass array hybridization

The SlideHyb Survey Kit contains four buffers that allow determination of the optimal hybridization parameters for a glass array system. The buffer that gives the greatest signal and the lowest background can then be purchased individually, providing customized glass array hybridization.

MicroCASTer

Schleicher & Schuell www.arraying.com

A handy device

Using this eight-pin manual arrayer, up to 768 spots can be placed on one slide. Processing time is 5–20 minutes per slide. An index system for 96-well source plates, which can be used to hold two slides for arraying at once, ensures correct positioning of the spots. The system is designed for use with CAST slides, which can be detected using radioactivity or chemiluminescence. Supplied complete with an eight-pin system hand tool and slide holder, pin conditioner and replicator pins, the device is especially suitable for labs without access to robotic arrayers and fluorescent scanners.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

CUSTOMISED monoclonal antibodies with GUARANTEE



BioGenes
Gesellschaft für Biopolymere mbH

Internet www.biogenes.de
Berlin *Germany
Phone +49 (0)30-65 76 23 96

BACs • YACs • PACs cDNAs

We provide clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

ResGen[™]
Invitrogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at
www.resgen.com

OAKTOWN PUBLICATIONS

100 Western Avenue, London W3 7TX
Tel: 020 8749 4555; Fax: 020 8740 7782

Two Seminal books by **Shafi U. Ahmed**
Fermat's Last Theorem: New Simple Proof
(also Goldbach), ISBN 094 8331 100, and
We Are Alone: The Limited Universe
(Proves speed of light is immeasurable)
ISBN 094 8331 151. Price £10 each plus
P&P £1(UK), £3(abroad).
Order through Booksellers or from us. Major
Credit/Debit Cards accepted. Key texts free on:
www.oaktownpublications.com

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: classified@nature.com

Group European Manager:

Leonie Weiss (4954)

European Manager:

Nevin Bayoumi (4978)

European Sales Consultant:

Jonathan Bull (4973)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Sille Opstrup (4994)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: classified@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Spain/ Portugal:

Amanda Wren

Tel + 34 91 391 3835

Fax + 34 91 319 0091

e-mail: a.wren@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: classified@nature.com

US Sales Director:

Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Cortisa Salzman

Japan Head Office, Tokyo

Shin-Mitsuke Building (4F),
3-6 Ichigaya Tamachi Shinjuku-Ku,
Tokyo 162
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@nature.com

Japan Manager: Kate Cowan

Warming up Scandinavia

During a cold, damp, grey March visit to the biotech nexus of Copenhagen, Malmo and Lund — also known as Medicon Valley — I was not surprised when my hosts told me that the weather, especially the long winters with their lack of both heat and light, gave pause to potential recruits from warmer climes. But despite these chilly conditions, the welcome that I was accorded could not have been warmer.

When in Scandinavia, I survived two days of interviews undertaken while I was suffering with flu. Indeed, normally, I would have cancelled my appointments — but they had been arranged months in advance, and I had just travelled thousands of miles to reach them. Fortunately, pharmaceutical executives, investment bankers, university researchers and biotech senior executives all endured patiently my fevered, disjointed questions — punctuated all-too-frequently by sneezes, coughs and me blowing my nose. Nonplussed, they proffered great quantities of tea, over-the-counter remedies and frank answers.

During the course of my queries, I learned that people on both sides of the Øresund Sound have worked to make Medicon Valley a better environment for scientists. New venture capital funds all but guarantee that any marketable biology born in the area has a chance to succeed. Biotech parks located close to universities in both cities encourage new relationships between academia and industry. And some business leaders say they have already made moves to attract the scientists from outside the area that the region will need if it is to grow (see story page 10).

Those efforts show collectively that even though you can't improve the weather, you can change the climate.

Paul Smaglik
Naturejobs editor



Contents

CAREERS AND RECRUITMENT

US academic health centres
fight trend for privatizing
clinical trials p4

Contract research organizations
offer jobs, but limit
publications p6

Doctor shortage in France
makes patient care a priority
over clinical research p7

REGIONS

Medicon Valley p10

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

From bench to bedside and back?

US academic health centres face challenges that may threaten the future vitality of the clinical research enterprise, but they are fighting back, says Diane Gershon.

At a time when public and private support for biomedical research in the United States is robust and the opportunities for biomedical advancement never greater, some of the nation's most prestigious research hospitals are financially ailing. These fiscal difficulties may be lessening the ability of the nation's 125 academic health centres (AHCs) to conduct clinical research and nurture the careers of the very cadre of researchers best positioned to translate basic research into clinical practice.

AHCs, which comprise medical schools and their affiliated teaching hospitals and faculty groups, straddle and integrate basic and clinical research. They provide an environment where basic and clinical researchers can rub shoulders and transfer what is learned at the bench to the bedside.

AHCs conduct nearly 30% of health-related research in the United States, according to a 1999 report, *From Bench to Bedside*, by the Commonwealth Fund Task Force on Academic Health Centers. In addition to research, AHCs also train health professionals and biomedical researchers, provide patient care and conduct community outreach.

The aspect of clinical research that seems most threatened is 'translational' clinical research, according to David Nathan, president emeritus of the Dana-Farber Cancer Institute in Boston and professor of paediatrics at Harvard Medical School. He defines translational research as studies in the lab that are brought directly to the patient to improve either diagnosis or therapy. And it is the cohort of clinical researchers doing translational research, with a

foot in the laboratory (and a reasonable grasp of basic research) and a foot by the bedside studying patients, that seem to be in short supply, he says.

The financial pressures facing some AHCs can be blamed, in part, on the advent of managed care in the United States during the 1990s. "It's not hard to show that these institutions [AHCs] are economically more pressed, at least on the clinical side, than they have been before," says David Blumenthal, director of the Institute for Health Policy at the Massachusetts General Hospital/Harvard Medical School. But to go from there and show that the central missions of the AHCs are compromised is more difficult, he says.

AHCs facing financial pressures have responded by cutting programmes, reducing workforces and selling teaching hospitals. Some AHCs are also creating a more industry-friendly environment in which to conduct clinical trials in the face of aggressive competition from for-profit entities (see page 6).

Having a stab: some academic health centres are creating a more industry-friendly environment in which to conduct clinical trials.



HARD ROAD TO HOE

Reduced clinical revenues from faculty group practices and affiliated teaching hospitals has meant less money in the pot for the cross-subsidy of clinical research and training. Moreover, in highly managed markets in particular, clinical faculty — especially junior faculty — are under increased pressure to devote more of their time to patient care and less to activities involving teaching and research.

Physicians contemplating a career in clinical research face endless training periods and enormous medical school debts. In 1999, the average medical school debt was about \$91,000, according to Stephen Heinig, senior staff associate in the Division of Biomedical and Health Sciences Research at the Association of American Medical Colleges (AAMC). Moreover, given concerns about the sustainability of such a career and the lure of higher incomes in clinical practice, it is not difficult to see why many physicians

turn their back on a career in clinical research.

Alarm bells first rang over the potential shortfall in clinical researchers over 20 years ago when James Wyngaarden spoke and wrote of the clinical investigator as an 'endangered species'. Clinical research was not helped when AHCs built up their capacity for basic research in the 1980s and 1990s. "I think AHCs are representative of the culture of science in the biomedical area and they have been infatuated with molecular biology for the past 20 years," says Blumenthal.

National Institutes of Health (NIH) grant applications from physician-scientists have remained roughly steady for the past 20 years, says Blumenthal. However, the implications of doubling the NIH budget may provide 'a strong argument' to increase the supply of people who could translate the growing body of research funded by NIH largesse into clinical practice, he says.

A BRIGHTER FUTURE?

AHCs may not be out of the woods financially, but the NIH's prosperity may help ease the situation. The agency has introduced several initiatives in the past few years aimed at creating a more hospitable environment for physicians pursuing a career in clinical research. These include the mentored, career development award (K23), designed to kick-start the careers of young investigators involved in patient-oriented research, the mid-career development award (K24) and the clinical research curriculum award (K30) given to institutions not individuals.

The K23 award is an "important recognition point and it provides [clinical researchers] with protected time," says Roger Meyer, the AAMC's senior consultant for clinical research in the Division of Biomedical and Health Services Research. The K23 award allows individuals to spend three-quarters of their time on research, away from patient-care duties, which is necessary if they are going to compete successfully for an NIH-supported RO1 grant, he says.

The University of Colorado Health Sciences Center in Denver, Colorado, was one of the lucky recipients of an institutional K30 curriculum award and for the past two years has offered both doctoral and certificate programmes in clinical science for the advanced training of physicians and healthcare professionals. "The NIH award mechanisms have truly come to our rescue," says Laurie Shroyer, associate director of the PhD Program in Clinical Science. The award was essential in getting the certificate programme underway, she says.

The clinical investigation track of the PhD programme, for example, has a strong mentoring component and builds proficiency in areas such as biostatistics, clinical epidemiology, clinical trial design, grant writing, biopharmaceuticals and pharmacokinetics. Although off to a good start, the uptake for these programmes, particularly among physician fellows, is not as uniform as Shroyer would like. Depending on their financial status, support for students varies across departments and divisions, she says.

The prospect of the introduction of an extramural, NIH-funded clinical research loan repayment programme (one already exists for intramural

researchers) is also energizing the community and would provide the welcome relief from educational debt that prevents so many from contemplating a clinical-research career. Congress authorized the programme late last year as part of the Clinical Research Enhancement Act and Pediatric Research Act. For every year of service as a clinical researcher (up to a maximum of three years), it would allow the NIH to repay up to \$35,000 of an individual's educational loans. the NIH must still devise an implementation strategy for the programme, which is unlikely to start before 2002 as no funds were appropriated for 2001.

CATCH 'EM YOUNG

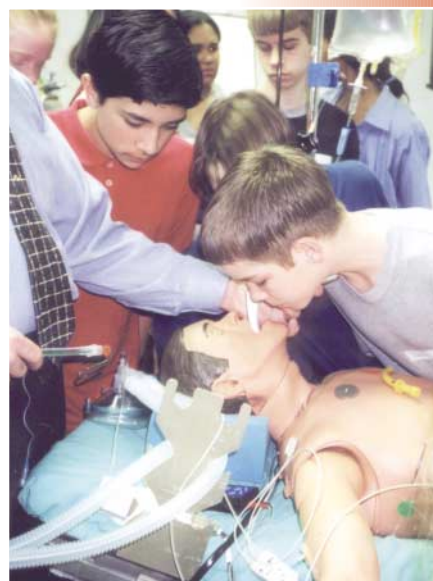
Some evidence indicates that individuals make career choices at an early age. With this in mind, institutions, such as Wake Forest University School of Medicine in Winston-Salem, North Carolina, now run professional development programmes for science and maths teachers in the community. The programmes at Wake Forest, for example, are undertaken in partnership with the local school system and the historically black Winston-Salem State University, and are geared towards closing the performance gap in maths and science education.

In addition to these community outreach initiatives, Wake Forest also strongly encourages students to gain research experience while in medical school. About one-quarter of the freshman-class students do a summer research rotation, the majority are in clinical research, and all medical students must complete an independent science-study before graduation.

Some of the foundations are also getting in on the act. The Doris Duke Charitable Foundation, for example, already offers two career-development awards aimed at strengthening and revitalizing clinical research. The foundation's newest programme will enable students to take a year off medical school and obtain some didactic and 'hands-on' clinical research training, as well as undertake a mentored clinical-research project. And the Damon-Runyon-Walter Winchell Foundation has received a \$15-million grant from Eli Lilly to help revive clinical investigation in cancer.

More support, from the NIH, foundations and the AHCs themselves along with a debt-repayment plan, may yet revive an ailing part of the biomedical research enterprise in the United States.

Diane Gershon is Assistant Editor, New Technology at *Nature Medicine*.



Putting your money where your mouth is: an NIH-funded programme offers to repay clinical education debts to encourage new trainees.

Some evidence indicates that individuals make career choices at an early age.



The movement of clinical trials from the public to the private sector is forcing academic health centres to re-examine their role.

Traditionally, drug companies both managed and sponsored clinical trials while academic health centres (AHCs) conducted them. But times are changing. Drug companies, under increasing pressure to find ways to cut costs, create efficiencies and reduce time-to-market, are increasingly opting to outsource the management of clinical trials to commercial entities. Many of the trials are now taking place in venues other than the AHCs, such as community hospitals, which are competing hard for a share of the clinical-trials market in much the same way they already do for patient care.

New for-profit players have entered the marketplace over the past two decades and include contract research organizations (CROs) and site-management organizations. There are even companies that match drug companies with CROs in an effort to streamline the outsourcing process. Although outsourcing of clinical trials is creating new jobs, a job in a CRO will not suit everybody and not everybody will thrive in such an environment. "It is a changing environment and so it's important to be able to cope with change and also be very solutions-driven," says Helen Wyn Davies, senior vice-president of clinical development services at Quintiles, the largest drug CRO in the United States. Moreover, as this is a service industry, "you're viewing life very much in terms of a provision of a service," she says.

The CRO industry draws a high proportion of its employees (about 80%, according to Joseph von Rickenbach, chairman and chief executive of Parexel, one of the top three drug CROs in the United States) from the pharmaceutical industry. There, it is not uncommon for individuals to focus on one therapeutic area and be assigned to a particular study for years at a time. One of the benefits — and challenges — of working within a CRO is that the companies offer the chance to work on several different therapeutic areas and studies, and with a range of clients, within a short space of time.

It was just such an opportunity that attracted Gary Gartenberg to Covance two-and-a-half years ago. He had been in private medical practice for 20 years before deciding on a career move. "It's been very motivational to be involved in a company that cuts across the entire industry," says Gartenberg, one of 15 physicians in the clinical-development team that designs and reviews the protocols and deals with the trial's medical aspects and safety monitoring.

Although working with

CROs provides variety, the job has disadvantages as well. For example, publication opportunities are fairly limited. "It doesn't happen as often as I would like," says Raul Valentin, vice-president of human resources for Covance's Clinical Development Group. As all CROs operate under strict confidentiality agreements and non-disclosure policies, clients must consent to publication of any information deemed proprietary.

REGAINING LOST GROUND

While some CROs continue to expand their global reach, and are moving into areas such as discovery, product development, manufacturing and sales, AHCs are rethinking their provision of clinical services. In an effort to create a more industry-friendly environment, some AHCs have formed centralized clinical-trial units to address the issues concerning patient recruitment times, delays in institutional review-board approval, quality assurance and higher costs that are causing drug companies to go elsewhere.

Some institutions have established formal relationships with CROs in the hope of expanding their access to industry-sponsored trials. Parexel's oldest and most successful alliance is with Humboldt University. "It's so intimate that our company has several hundred employees on campus in Berlin," says von Rickenbach, and physicians at the university can intern with Parexel. von Rickenbach admits, though, that forging such partnerships can be arduous work. "We are a commercial company... and academic centres are not. Their primary purpose is not to earn a return to shareholders and to get these two purposes... onto a common denominator is not easy," he says.

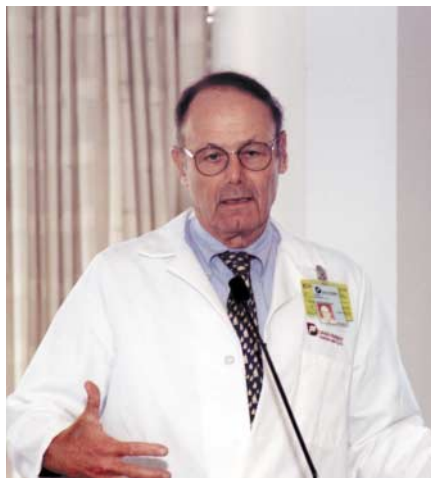
Indeed, not all such partnerships prove fruitful. Three years ago, Johns Hopkins Medical Institutions entered into a non-exclusive agreement with Quintiles. Johns Hopkins hoped that Quintiles would steer more clinical studies its way and in return Quintiles would have the name recognition of Hopkins and access to its top clinical investigators. "In reality [Quintiles] didn't give us anything and they weren't very interested I think in developing a real lasting relationship," says endocrinologist Adrian Dobs, at Johns Hopkins. The partnership has since fizzled out.

CROs generally feel that AHCs are basically too slow for them, says Dobs. "We appreciate that and we are trying our best to do better but we're under different constraints," she says. Johns Hopkins established a clinical-trials unit about two years ago. The unit helps sponsors identify investigators but also offers investigators a way to improve how they market their expertise. But as the line between publicly funded and privately sponsored research blurs, issues of academic freedom and conflict of interest need to be worked out more carefully, says David Nathan, president emeritus of the Dana-Farber Cancer Institute in Boston and a paediatrician at Harvard Medical School.

"If universities get involved with it for the money, I worry about that too because it's not really the academic mission," he says. "The science is moving like mad. It's extremely exciting and I think that's where the AHCs should be, not running big clinical trials."

D. G.

Paediatrician David Nathan believes the AHCs should be at the cutting edge of research, not running big clinical trials.



Too few doctors in the house

Increased demand by public healthcare for an already dwindling number of doctors has resulted in a low supply of clinical researchers in France, says Catherine Tastemain.

With the falling number of graduate doctors, the growing lack of qualified 'back-up' staff and a university hospital system hardly conducive to the setup of clinical trials, French public clinical research is in a predicament. "Any young doctor who wants to pursue clinical research certainly has his work cut out for him," says Christian Funck-Brentano, a clinical pharmacologist at the St Antoine university hospital centre in Paris, who also heads the hospital's clinical investigation centre (CIC). His colleagues heartily agree. Young graduates have to overcome countless hurdles if they want to pursue clinical research in a public hospital and solutions do not seem to be forthcoming despite the incentives that the authorities have deployed over the past few years.

Until 1970, France was churning out roughly 10,000 physicians a year. But in the early 1980s, a competitive end-of-year examination was enforced for first-year medical students. That test has greatly reduced the supply of potential physicians. Since 1985, a mere 3,500 doctors have been trained per year. "Today's predicament is that there are just not enough of us to care for the patients," pointed out Marc Beaussier, a young anaesthetist at St Antoine hospital.

So, as a 'hospital practitioner', he is having an even harder time undertaking clinical research. In France, hospital practitioners, who only receive their stipend from the hospital, are 'encouraged' to do clinical research even though their professional status does not include research. When Beaussier wants to start a clinical investigation, he must take time off his medical practice at the hospital.

However, protocols take time to put together, administrative delays take forever and the qualified staff that could help him design the project, collect the data and analyse research findings are lacking. "It's very long and very complicated and nobody will pay the hospital back for the time I'm going to spend doing all this myself," says Beaussier.

Actually, hospital practitioners who want to undertake clinical research do have a solution — to become a senior lecturer or university professor combining the university and the hospital. If they succeed, they are required to devote 50% of their time to teaching and research. In practice the figure is usually 20% or even 30% because of healthcare needs.

Also, universities emphasize some kinds of research over others, says Jean-Claude Dussaule, a physiologist at St Antoine. "Clinical research is easy to get started when it addresses a drug because trials are heavily backed by the pharmaceutical industry," he says. "On the other hand it's all uphill for clinical research on physiopathology."

FRENCH PHYSICIANS FLEE PHARMA

From 1980 to 1995, many highly qualified public-sector doctors turned to private pharmaceutical R&D because of higher salaries and the attraction of responsible positions. The stream of physicians flowing into this private industry has now dwindled. Current job offers are not considered very gratifying for doctors and, above all, hospital demand is now high because the country has fewer doctors. But the opportunities for young scientists are still attractive.

Emmanuelle Le Gorfec, a young biostatistician in her third year working for a private pharmaceutical company, said: "Here, the pay and recognition are better whereas at the university or public research organizations

Joël Ménart, head of clinical research, Paris Public Hospitals, believes that poor organization, not lack of funds, is the major obstacle to clinical research in hospitals.



you're just one in a thousand."

The pharmaceutical industry has, of course, maintained its close ties with public hospitals, especially university-hospital facilities, with the industry funding a large share of the clinical trials of drugs. Increasingly, contract research organizations (CROs, see p.6) are standardizing operational protocols. Hospital clinicians generally admit that, overall, CROs with more resources on tap than public hospitals do their job well. However, a trend has surfaced over the past several years. Although the pharmaceutical industry has kept clinical trial centres in Western Europe (because the drugs are mainly sold in these countries), the industry is organizing more clinical studies in the former Eastern Bloc.

HOSPITALS SOLD SHORT

In contrast to this excess of riches in CROs, Joël Ménart, the top official in charge of the Clinical Research Division, Paris Public Hospitals, notes that clinical research at hospitals tends to get short shrift. The problem, he says, is that researchers at university hospitals must divide their time among teaching, care, research and administration simultaneously. Ménart thinks that, instead, clinicians should rotate from duty to duty in shifts, rather than juggle them all simultaneously. "I'm not saying it's easy to organize. It would probably take several years to set up," says Ménart.

Although Ménart's proposed reform is not yet being considered, other changes have been gradually introduced over the past several years, but to limited effect. For example, the government decided to set up CICs in some hospitals to bring together hospital departments and research units from Inserm, the

French national institute for health and medical research.

Although far-reaching, the changes have not been enough to turn the situation around. The 17 CICs now in place cannot realistically be considered as providing enough outlets for young physicians. Many doctors — even the ones working at a university-hospital centre — feel that there is a conflict between hospital budget constraints and the cost of clinical research. "The authorities can't tell us that the prescribed antibiotics are too expensive out of one side of their mouth and announce a several million-franc investment in clinical research out of the other," said Beaussier.

Ménart disagrees. The clinical-investigation budget is negligible compared to the budget allocated for care. "If you tap into every possible funding source, you get an overall appropriation of roughly 150 million francs (US\$19.5 million) — not including the wages for 3,000 university-hospital teachers — whereas health expenditures are at a hefty 696 billion." He believes that in any case the budget is not the primary hurdle for clinical research in hospital but rather the poor organization of the university-hospital system. ■

Catherine Tastemain is a science journalist based in Paris.

UK WORKING FOR A BETTER CLINICAL CLIMATE

Rarely have areas of basic science been more ripe for development into applications than those of the molecular genetics and biology flowing from the human genome project. Yet there is a

global lack of the clinical research pivotal to the transformation. "Less is being done, and it is less well done," says Helen Cope of the Clinical Careers Initiative at the Wellcome Trust.

In an attempt to reverse this trend in Britain, the Trust has for the past three and a half years been funding the building or refurbishment and equipping of clinical research facilities in Birmingham, Cambridge, Edinburgh, Manchester and Southampton. The National Health Service Trusts will pay the running costs, including salaries for core staff, such as research

managers and nurses, while the research councils will pay for the research through grants to clinical researchers at the universities.

The aim of the initiative is to bring the clinical and academic worlds closer together, something that is essential, says Cope, if the science emerging from the human genome project is to reach patients. Yet the facilities alone will not be enough. Doctors pursuing serious research usually take longer to reach consultant level than if they followed a service role within the NHS. Both the Wellcome and the Medical Research Council attempt to make

the research option more attractive by offering fellowships.

Several diploma and masters courses in the United Kingdom give an overview of drug development for anyone working in clinical research or some other aspect of drug development within any clinical organization.

If a drug is to be licensed for use, each step of the process must also satisfy the regulatory authorities. "Regulatory affairs is a growth area and will continue to be," says Nicky Lillioft, of the Association of British Pharmaceutical Companies. **Helen Gavaghan**

♦ <http://www.abpi.org.uk>

♦ <http://www.bira.org.uk>

For the past three years, the Wellcome trust has been funding the building and equipping of clinical research facilities in several UK universities.

WELLCOME TRUST





Bridging sectors Medicon Valley

A bridge spanning the Øresund Sound now connects biotech companies in southern Sweden and northern Denmark. Associations with local universities tie the companies to the area, and a plethora of venture capital funds fuel them. But for Medicon Valley's growing body of biotech companies to take off, they need one more resource that is in increasingly short supply — scientists.

"The bottleneck used to be venture capital," says Anker Lundemose, chief executive of Pantheco, a Danish biotech company. "The bottleneck now is human resources." About 2,000 R&D employees work in the region's 96 companies. But the growth in both new companies and venture funds means that more scientists will be needed.

The area must attract more talent from beyond what has traditionally been a fairly insular community, say area bankers, biotech executives and university administrators. And it will also need to retain the best and brightest PhDs produced by the concentration of universities that created the Medicon Valley Academy — the organization that basically organized the region's biotech into one operational unit (see *Nature* **395**, 412–413; 1999).

Claus Bræstrup, executive vice-president, R&D, of Lundbeck, the Copenhagen-based pharmaceutical firm, says companies will need to

US satellite companies

Medicon Valley is attracting US companies, meaning more jobs — but not necessarily in R&D. The latest and largest catch is Biogen, a Cambridge, Massachusetts firm best known for interferon- β -1a.

Last month Biogen announced that it would move to a 60-acre site in Hillerød to establish a European manufacturing plant, expected to open in 2005, and to employ about 400.

Peter Iversen, programme manager for CIPHERGEN's Copenhagen location,

says that US-based companies that open satellite locations in Medicon Valley but leave their R&D efforts back home offer jobs for people not interested in basic research.

The Copenhagen company uses its five staff scientists to train customers how to use its protein chip, which purifies proteins and amplifies the signals read by mass spectrometers. But the company this summer will open a proteomics discovery centre in Copenhagen, which will create technical jobs in contract research.

P. S.

Pointing the finger: Pantheco chief executive Anker Lundemose pinpoints the shortage of scientists as the major bottleneck to biotech growth.



look outward more but successful overtures to researchers in other countries may not be easy. "There are major obstacles for people coming in to Denmark," Bræstrup says. "One is the taxes; another is the weather."

Denmark has already begun to address the tax

Danish and Swedish cooperation

The Medicon Valley Academy consortium has helped foster cooperation in both Denmark and Sweden, allowing each country to share scarce resources. In Sweden, members of the consortium formed SWEGENE, a postgenomics research organization, and in Denmark the Biotech Research and Innovation Centre (BRIC), a similar postgenomics centre, is being established.

SWEGENE was initiated in 2000 to promote research in functional genomics. The consortium, which encourages collaboration between different universities, is funded by the Knut and Alice Wallenberg Foundation with contributions from the collaborating universities.

SWEGENE, rather than supporting individual scientists or research groups, is funding shared resource

centres in bioinformatics, structural biology, phenotype profiling and genetic variability. Funds so far have been awarded to develop a sequencing centre, a gene-expression centre and to support postdocs.

BRIC, the Danish equivalent, is at an earlier stage of development and will rely more on government grants, participating institutions and industry for funding. BRIC is helping to fund a 27,000 m² research facility it will share with the University of Copenhagen and the Copenhagen Hospital Cooperative. BRIC will employ about 150 scientists and technicians in the new facility in 2004. P. S.

Web links

SWEGENE
<http://www.swegene.org>
 Biotech and Innovation Research Centre
<http://www.bric.dk>

system, which levies rates as high as 62% for the best-paid scientists. Last year, it established an 'expert' bracket, which allows out-of-country workers with skills in demand to pay only 25% for three years.

The country had a similar programme before, but it only applied to the highest-paid positions — and anyone deciding to stay on after three years was required to pay back the difference. More people are eligible under the new system, and the assimilation penalty no longer applies. But the tax break can cause resentment among co-workers who pay the higher rates, Bræstrup says.

Lundbeck has recruited extensively outside the country — adding 117 R&D workers in the past year and doubling its research budget in the past two. The company is especially short on organic chemists and bioinformatics experts.

Kirsten Drejer, chief executive of Symphogen, a Danish biotech company specializing in developing human antibodies to viruses, says she is having an especially hard time finding senior people. She agrees with Bræstrup that the Scandinavian weather — especially the long, dark winter — can make it hard to recruit scientists from warm sunny regions such as California. "On the other hand, it's very safe," says

Søren Mouritsen, chief executive of M&E Biotech, one of the first Danish biotech companies. He tries to emphasize to recruits that the area is considered to offer excellent quality of life.

RETAINING TALENT

Meanwhile, some scientists who were trained in the area are reluctant to wholeheartedly embrace industry. "The public researcher is still a little bit shy about collaborating with private investors," says Hans Poulsen, chief executive of Odin Medical, a Copenhagen-based company that is developing nonviral gene-therapy vectors to fight cancer.

Poulsen has combated that cultural tendency by making his company into a bit of a hybrid. It is supported in part by the Danish Cancer Institute and is housed on the University of Copenhagen campus. Poulsen also expects his employees to do basic research and publish their findings. As a result of this approach, he has been successful in finding qualified candidates. It is important to offer a research setting in a non-academic environment, because academic slots have been increasingly scarce due to a dearth of government funding, Poulsen says.

Eva Degerman, from Lund University's section for molecular signalling, is concerned about a similar trend in Sweden. She would prefer not to have to rely on industry to fund her research. "I think it is important that we have our independence," she explains.

But that may become increasingly difficult. Degerman, like many Swedish scientists, depends on government and foundation money to support her work. These funds have not been rising nearly as fast as they have in the United States, and she is worried that too much dependence on industry money will result in less freedom to publish and an over-emphasis on product over basic research.

Per Belfrage, former dean of Lund's medical school, thinks that Medicon Valley-area academic scientists who learn to interact with the region's industry can help themselves retain their independence.

For example, if industrial and academic scientists band together under the aegis of the Medicon Valley Academy, they may be able to lobby successfully for more basic biomedical funding — something scientists from both sectors say is lacking. Although the government cannot control the climate, they can do something about funding. ■

Paul Smaglik is Naturejobs editor.

Web links

Medicon Valley Academy
<http://www.mva.org>
 Copenhagen Capacity
<http://www.copcap.com>



Cold comfort: the long, dark Scandinavian winter can deter scientists from warmer regions, says Kirsten Drejer, chief executive of Symphogen.



Independent streak: Lund University's Eva Degerman worries that reliance on industry for funding will result in less academic freedom.